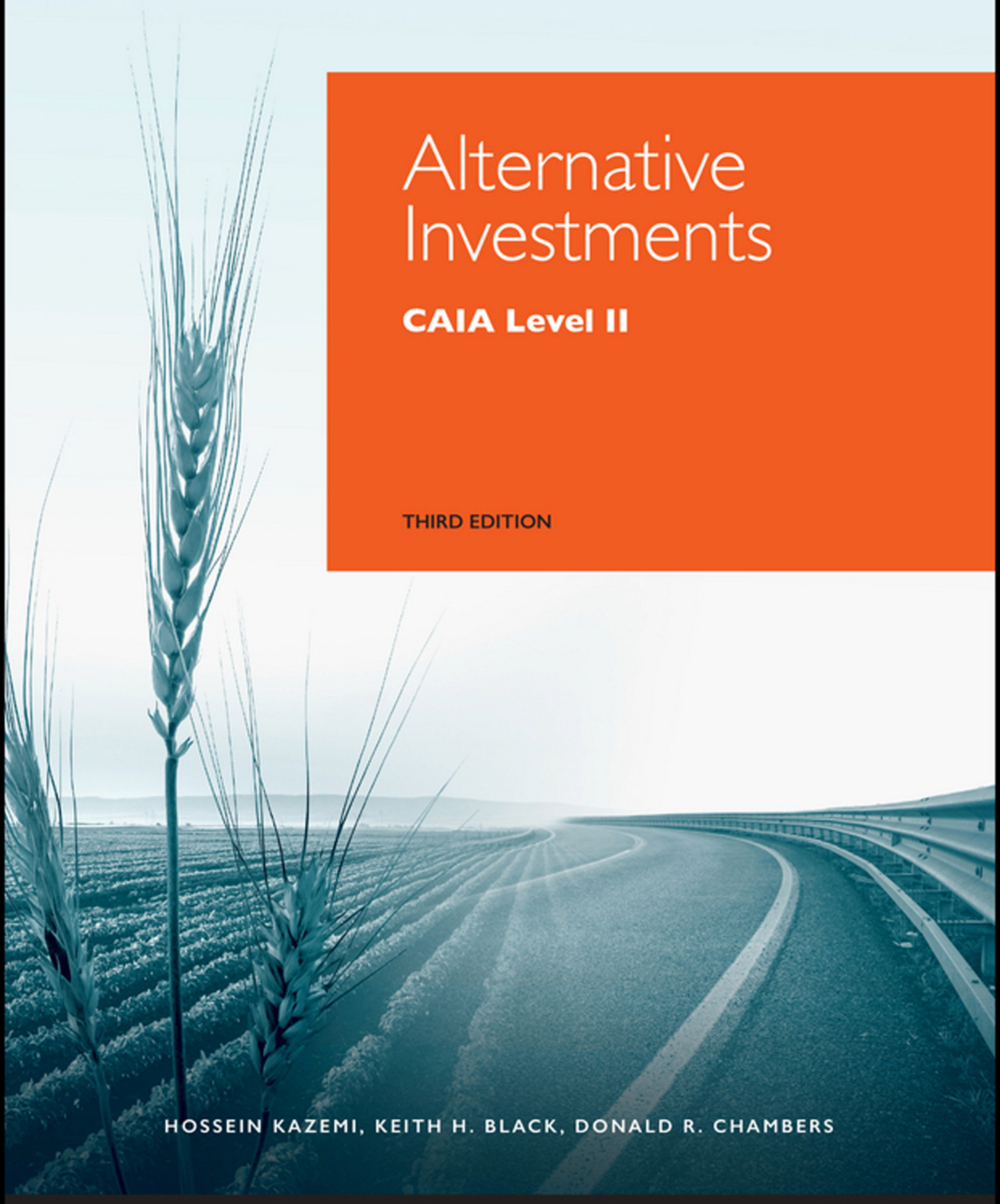
CAIA ASSOCIATION®

CAIA KNOWLEDGE SERIES

Alternative Investments

CAIA Level II

THIRD EDITION



HOSSEIN KAZEMI, KEITH H. BLACK, DONALD R. CHAMBERS

Alternative Investments

The Wiley Finance series contains books written specifically for finance and investment professionals as well as sophisticated individual investors and their financial advisors. Book topics range from portfolio management to e-commerce, risk management, financial engineering, valuation and financial instrument analysis, as well as much more. For a list of available titles, visit our website at www.WileyFinance.com.

Founded in 1807, Wiley is the oldest independent publishing company in the United States. With offices in North America, Europe, Australia and Asia, Wiley is globally committed to developing and marketing print and electronic products and services for our customers' professional and personal knowledge and understanding.

Alternative Investments

CAIA Level II

Third Edition

HOSSEIN B. KAZEMI
KEITH H. BLACK
DONALD R. CHAMBERS

WILEY

Cover image: © iStock & Shutterstock
Cover design: Zoe Design Works

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

ISBN 9781119016397 (Hardcover)
ISBN 9781119016366 (ePDF)
ISBN 9781119016380 (ePub)

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

PART 2
Private Equity

PART 3
Real Assets

PART 6
Structured Products

Preface

Alternative Investments: CAIA Level II is designed as the primary reading resource for the Level II exam of the Chartered Alternative Investment Analyst (CAIA) Association's Charter program. To ensure that the material best reflects up-to-date practices in the area of alternative investments, the CAIA Association invited a group of leading industry professionals and academics to contribute to the production of this book. While some of them helped directly by writing some of the chapters of this book, others provided valuable input as members of our advisory board. Without their immense talent and dedication, this book would not have been completed.

Since its inception in 2002, the CAIA Association has strived to be the leader in alternative investment education worldwide and to be the catalyst for the best education in the field wherever it lies. The CAIA program was established with the help of a core group of faculty and industry experts who were associated with the Center for International Securities and Derivatives Markets (CISDM) at the Isenberg School of Management and the Alternative Investment Management Association (AIMA). From the beginning, the CAIA Association recognized that a meaningful portion of its curriculum must be devoted to codes of conduct and ethical behavior in the investment profession. To this end, with the permission and cooperation of the CFA Institute, we have incorporated its Code of Ethics and its *Standards of Practice Handbook* into our curriculum. Further, we have leveraged the experience and contributions of our members and other alternative investment professionals who serve on our board and committees to create and update the CAIA Association program's curriculum and its associated readings.

The quality, rigor, and relevance of our curriculum readings derive from the ideals upon which the CAIA Association was based. The CAIA program offered its first Level I examination in February 2003. Our first class consisted of 43 dedicated investment professionals who passed the Levels I and II exams and met the other requirements of membership. Many of these founding members were instrumental in establishing the CAIA designation as the global mark of excellence in alternative investment education. Through their support and with the help of the founding cosponsors—the AIMA and the CISDM—the CAIA Association is now firmly established as the most comprehensive and credible designation in the rapidly growing sphere of alternative investments.

The AIMA is the hedge fund industry's global, not-for-profit trade association, with more than 1,500 corporate members worldwide. Members include leading hedge fund managers, fund of hedge funds managers, prime brokers, legal and accounting services, and fund administrators, all of whom benefit from the AIMA's active influence in policy development; its leadership in industry initiatives, including education and sound practice manuals; and its excellent reputation with regulators.

The CISDM of the Isenberg School of Management at the University of Massachusetts–Amherst seeks to enhance the understanding of the field of alternative investments through research, education, and networking opportunities for member donors, industry professionals, and academics.

The CAIA Association has experienced rapid growth in its membership over the past 14 years. It is now a truly global professional organization, with more than 8,000 members in over 80 countries. We strive to stay nimble in our process so that the curriculum remains relevant and keeps pace with the constant changes in this dynamic industry.

Although the CAIA Association's origins are largely based in the efforts of professionals in the hedge fund and managed futures space, these founders correctly identified a void in the wider understanding of alternative investments as a whole. From the beginning, the CAIA curriculum has also covered private equity, commodities, and real assets, always with an eye toward shifts in the industry. Today, several hundred CAIA members identify their main area of expertise as real estate or private equity, and several hundred more are from family offices, pension funds, endowments, and sovereign wealth funds that allocate across multiple classes within the alternative investment industry. To ensure benefit to the widest spectrum of members, we have developed curriculum subcommittees that represent each area of coverage within the curriculum. Alternative investment areas and products share some distinct features, such as the relative freedom on the part of investment managers to act in the best interests of their investors, alignment of interests between asset owners and asset managers, and relative illiquidity of the investment positions of some investment products. These characteristics necessitate conceptual and actual modifications to the standard investment performance analysis and decision-making paradigms.

Our curriculum readings are designed with two goals in mind: first, to provide readers with the tools needed to solve problems they encounter in performing their professional duties; and second, to provide them with a conceptual framework that is essential for investment professionals who strive to keep up with new developments in the alternative investment industry.

Readers will find the publications in our series to be beneficial, whether from the standpoint of allocating to new asset classes and strategies in order to gain broader diversification or from the standpoint of a specialist needing to better understand the competing options available to sophisticated investors globally. In both cases, readers will be better equipped to serve their clients' needs.

CAIA Level II required readings consist of three parts: this book and the CFA Institute's *Standards of Practice Handbook* and *Current and Integrated Topics Readings*. Information about obtaining the last two components can be found on our website, caia.org. Many resources are freely available on our website as well.

We will continue to update the *CAIA Level II Study Guide* every six months (each exam cycle). The study guide outlines all of the readings and corresponding learning objectives (LOs) that candidates are responsible for meeting. The guide also contains important information for candidates regarding the use of LOs, testing policies, topic weightings, where to find and report errata, and much more. The entire exam process is outlined in the *CAIA Candidate Handbook*, which is available at caia.org. Candidates can also access a workbook that solves the problems presented at the end of each chapter and other important study aids.

We believe you will find this series to be the most comprehensive, rigorous, and globally relevant source of educational material available within the field of alternative investments.

Hossein Kazemi, PhD
Senior Adviser to the CAIA Association

Acknowledgments

We would like to thank the many individuals who played important roles in producing this book. In particular, we owe great thanks to William Kelly, Chief Executive Officer of the CAIA Association, and our committee members:

Curriculum Advisory Council

Stephane Amara, CAIA
Mark Anson, CAIA
Garry Crowder
David McCarthy
Tom Robinson, CAIA
Hilary Till
James Tomeo

Hedge Funds, CTAs, and Fund of Hedge Funds Committee

Jaeson Dubrovay, CAIA
Mark Hutchinson
Kathryn Kaminski, CAIA
Jim Liew
Hamlin Lovell, CAIA
Putri Pascualy
Mark Wiltshire, CAIA

Real Assets (Real Estate, Commodities, Infrastructure, Intellectual Properties, and Natural Resources) Committee

Tom Arnold, CAIA
Andrew Baum
Georg Inderst
Sameer Jain
Tom Johnson, CAIA
David Lynn
George A. Martin

Joelle Miffre
Richard Spurgin

Private Equity and Venture Capital Committee

James Bachman, CAIA
Erik Benrud, CAIA
Douglas Cumming
Ludovic Phalippou
Pierre-Yves Mathonet
Thomas Meyer
Gitanjali M. Swamy

Due Diligence, Risk Management, and Regulation Committee

Gordon Barnes, CAIA
Michal Crowder
Jason Scharfman
Christopher Schelling, CAIA
Sean Gill, CAIA
Tom Kehoe
Danny Santiago, CAIA

Asset Allocation, Endowments, Pension Funds, and Sovereign Funds Committee

Samuel Gallo, CAIA
James T. Gillies, CAIA

Special credit goes to CAIA staff for their valuable contributions in painstakingly bringing the third edition to completion.

CAIA Staff

Stephen Abernathy, Associate Director of Research and Publications
Nelson Lacey, Director of Exams
Kathy Champagne, Senior Associate Director Exams Administration
Kristaps Licis, Senior Associate Director of Exams
Nancy E. Perry, Publications Coordinator

Outside Editor

Jamie Thaman

About the Authors

Hossein Kazemi received his PhD from the University of Michigan, Ann Arbor and is a senior adviser to the CAIA Association. He is the Michael and Cheryl Philipp Professor of Finance at the University of Massachusetts–Amherst, Director of the Center for International Securities and Derivatives Markets, a cofounder of the CAIA Association, and Editor-in-Chief of *The Journal of Alternative Investments*—the official publication of the CAIA Association. He was a managing partner at Schneeweis Partners and Alternative Investment Analytics. He has authored or coauthored more than 30 scholarly articles and is a coauthor of *The New Science of Asset Allocation: Risk Management in a Multi-Asset World* (2010, John Wiley & Sons) and *Postmodern Investment* (2013, John Wiley & Sons).

Keith Black received his PhD at the Illinois Institute of Technology, Chicago. He serves as Managing Director of Curriculum and Exams at the CAIA Association. He was previously an Associate at Ennis Knupp and an Assistant Professor at Illinois Institute of Technology. He is a member of the editorial board of *The Journal of Alternative Investments*. He is also a CFA Charter Holder and a member of the inaugural class of CAIA candidates. He is the author of *Managing a Hedge Fund* (2004, McGraw-Hill). He was named to *Institutional Investor* magazine's list of "Rising Stars of Hedge Funds" in 2010.

Don Chambers received his PhD from the University of North Carolina, Chapel Hill. He is Associate Director of Programs at the CAIA Association; the Walter E. Hanson/KPMG Professor of Finance at Lafayette College in Easton, Pennsylvania; and Chief Investment Officer of Biltmore Capital Advisors. Professor Chambers previously served as Director of Alternative Investments at Karpus Investment Management. He is a member of the editorial board of *The Journal of Alternative Investments*. He is also a CAIA Charter Holder and the primary author of *Alternative Investments: CAIA Level I,* third edition (2015, John Wiley & Sons).

Mark Anson is Chief Investment Officer of Commonfund. He is responsible for overall client asset allocation, portfolio management, manager research and due diligence across equities, fixed income, and hedge funds. Prior to joining Commonfund, he was Chief Investment Officer for the Bass Family Office. Previously, Mark has served as President of Nuveen Investments, Chief Executive Officer and Chief Investment Officer for Hermes Pension Management and for the British Telecom Pension Scheme, and the Chief Investment Officer for CalPERS. Mark currently serves on the Executive Advisory Board of MSCI-Barra, The Investment Advisory Council of the UAW Pension Fund, the Law Board of Northwestern University School of Law, and the Board of Directors for the Chartered Alternative Investment Association. Mark earned a BA in Economics and Chemistry from St. Olaf College, a JD from

Northwestern University School of Law, and a PhD and Master's in Finance from Columbia University Graduate School of Business. Mark earned the Chartered Financial Analyst, Chartered Alternative Investment Analyst, Certified Public Accountant, and Chartered Global Management Accountant professional degrees, and is a Member of the Bar of the State of New York and the State of Illinois.

Jim Campasano is the President of Marshall James Capital, LLC, an advisory firm focusing on volatility products. A graduate of Harvard University with a degree in Economics, cum laude, he received a JD from Vanderbilt University School of Law and is a PhD candidate in finance at the Isenberg School of Management, University of Massachusetts–Amherst. Prior to Marshall James Capital, Mr. Campasano worked as a portfolio manager at Vicis Capital and Millennium Limited Partners, where he ran a long volatility, cross-asset portfolio. He contributed to Chapter 30 (Volatility, Correlation, and Dispersion Products and Strategies).

Michal E. Crowder received her JD from Northwestern University School of Law in Chicago and has a Master of Arts in Political Science from Northwestern University. Ms. Crowder has worked for several hedge fund and investment management firms over the past eight years and has traveled extensively throughout Europe and Asia. She is fluent in four languages and supports a number of not-for-profit endeavors. Ms. Crowder is licensed to practice law in Illinois and currently clerks for the Honorable Judge Abdul Kallon in the United States District Court of Northern Alabama. She is the primary author of Chapter 34 (Regulation and Compliance).

Satyabrota Das has more than 10 years of experience working in financial markets. He has developed and traded hedge fund and CTA replication products using liquid exchange-traded securities. Most recently, he developed an interactive web-based replication program that allows investors to create customized replication portfolios. Previously, he supported the Alternative Commodity Benchmark Index, a second-generation commodity index, for Alternative Investment Analytics, LLC. He is a CFA and CAIA Charter Holder, and is working on his PhD at the Isenberg School of Management, University of Massachusetts–Amherst. He is the primary author of Chapter 31 (Hedge Fund Replication).

Malay K. Dey is currently a senior partner of FINQ LLC, a diversified financial technology startup. He held faculty positions at the University of Illinois at Urbana Champaign, Cornell University, and the University of Minnesota, Twin Cities. Professor Dey has frequently visited the Indian Institute of Management Calcutta (IIMC) and has lectured at ISI Calcutta and other leading Indian institutions. He was a Research Fellow at the Networks Financial Institute at Indiana State University (2006–2008) and served as a Vice President, quantitative trading strategy, at ITG from 2006 to 2007. Professor Dey received his PhD in Finance from the Isenberg School of Management, University of Massachusetts–Amherst. His research focuses on theoretical and empirical issues related to institutional trading and liquidity in equity markets. He contributed to Chapter 27 (Relative Value Strategies).

Jaeson Dubrovay is a Managing Director at Blackcomb Holdings, Inc., an independent investment company. Previously he was a partner and cohead of Americas advisory for Aksia, LLC, one of the largest hedge fund specialty consulting firms. Prior to that, he was the Senior Strategist, Hedge Funds, at NEPC LLC, one of the industry's leading general investment consulting firms. Mr. Dubrovay has been managing money and consulting with leading institutional investors in connection with their hedge fund portfolios for more than 25 years. In 2008, he was named the Hedge Fund Consultant of the Year (*Institutional Investor*) and recognized for his contribution to the Investors Committee of the President's Working Group on Financial Markets, on Hedge Fund Best Practices. In 2009, Mr. Dubrovay was named Consultant of the Year by Foundation & Endowment Money Management (*Institutional Investor*) and was the major contributor to the team at NEPC that was named *PLANSPONSOR* magazine's Alternative Asset Consultant of the Year. He is a CPA and CAIA Charter Holder. He holds an MBA with honors from Santa Clara University. He is the primary author of Chapter 32 (Funds of Hedge Funds and Multistrategy Funds).

Urbi Garay received a PhD in Finance from the Isenberg School of Management, University of Massachusetts–Amherst, an MA from Yale University, and a BA in Economics from *Universidad Católica Andrés Bello* (Caracas, Venezuela). He is a Professor of Finance at the IESA Business School (Caracas, Venezuela). He was a visiting researcher at the CISDM (2007–2008), and has been a visiting professor at various business schools in Latin America, the United States, and Europe. He has been a consultant to the Inter-American Development Bank, the Venezuelan Central Bank, and the Caracas Stock Exchange. He is a coauthor of *Fundamentals of Finance* (IESA, 2005) and Long Term Investing (IESA, 2007). He has published articles in *The Journal of Alternative Investments*, *Emerging Markets Review*, *Emerging Markets Finance and Trade*, *Econometrics*, *Corporate Governance: An International Review*, and the *Journal of Business Research*. He is the primary author of Chapters 14–18 (Real Estate) and 35–36 (Structured Products).

Kathryn Kaminski is a Director at Investment Strategies at Campbell & Company. Prior to her recent move to Campbell & Company, she was Deputy Managing Director at the Institute for Financial Research (SIFR) and affiliated faculty at the Stockholm School of Economics. She is a featured contributor to the CME Group. Kathryn has experience working for a CTA fund of funds as well as quant experience in both emerging fixed income and credit markets. She lectures on derivatives, hedge funds, and financial management at the Stockholm School of Economics and has lectured previously at the Swedish Royal Institute of Technology (KTH) and the MIT Sloan School of Management. Kathryn completed her PhD at MIT Sloan, conducting research on financial heuristics. Kathryn is a coauthor of *Trend Following with Managed Futures: The Search for Crisis Alpha* (2014, John Wiley & Sons). Kathryn is a 100-Women in Hedge Funds PAAMCO CAIA Scholar and a CAIA Charter Holder. She is the primary author of Chapters 25 and 26 (Managed Futures).

Jim Kyung-Soo Liew is an Assistant Professor of Finance at Johns Hopkins Carey Business School. Dr. Liew teaches Advanced Hedge Fund Strategies, Corporate Finance, Derivatives, Entrepreneurial Finance, Fixed Income, and Wealth Management at the Johns Hopkins Carey Business School. Prior joining Johns Hopkins, Dr. Liew taught Statistical Arbitrage at Columbia University and CUNY Baruch College, and Hedge Fund Strategies at NYU Stern School of Business, as an Adjunct Professor.

Prior to that, he worked in the hedge fund industry where he built and implemented systematic investment strategies. Dr. Liew currently serves on the Editorial Advisory Board of *The Journal of Portfolio Management*. He resides with his wife and two daughters just outside of Baltimore. He is the primary author of Chapter 28 (Hedge Funds: Directional Strategies).

George Martin is a Senior Advisor to Wood Creek Capital Management, a real assets investment manager that is an affiliate of MassMutual and its asset management subsidiary Babson Capital Management. At Wood Creek, he has particular responsibility for matters related to research, portfolio construction, and risk management, and with a sector focus that emphasizes mid- and upstream agriculture. He is also a Senior Research Associate for the Center for International Securities and Derivatives Markets (CISDM) at the University of Massachusetts, Amherst, and a member of the Editorial Board of *The Journal of Alternative Investments*. He has been commercially active in real asset investing and commodity-based investments for the past decade. He is regularly called upon to speak on various aspects of the Alternative Investment business, and frequently publishes his research. Previously, he was a Research Fellow at the Brookings Institution. He has a BA and MA from Johns Hopkins University. He is the primary author of Chapters 20 and 21 (Real Assets).

Pierre-Yves Mathonet is Head of Risk in the Private Equities Department of the Abu Dhabi Investment Authority. He is a permanent member of the EVCA's Risk Measurement Guidelines working group. He codirected the Certificate in Institutional Private Equity Investing (CIPEI) course held by the Oxford Saïd Business School's Private Equity Institute. Previously, he was the head of the private equity risk management division of the European Investment Fund (part of the European Investment Bank group), worked as an investment banker in the technology groups of Donaldson, Lufkin & Jenrette and Credit Suisse First Boston, and, earlier, for the audit and consulting departments of PricewaterhouseCoopers. Pierre-Yves has coauthored several books including *Beyond the J Curve* (2005, John Wiley & Sons) and *J Curve Exposure* (2007, John Wiley & Sons). He holds a Master of Science cum laude in Finance from London Business School and a Master of Science magna cum laude in Management from Solvay Business School in Brussels. He is also a Certified European Financial Analyst cum laude. Pierre-Yves Mathonet and Thomas Meyer are the primary authors of Chapters 7–9 and 11–13 (Private Equity and Venture Capital).

Thomas Meyer is partner and cofounder of LDS Partners, specializing in the development of investment strategies, portfolio management, cash-flow forecasting, and asset allocation models for real assets (private equity, infrastructure, real estate). Mr. Meyer was responsible for the creation of the European Investment Fund's risk management function and was a director of EVCA (now Invest Europe). He was the secretary of the EVCA Private Equity Risk Measurement Group, codirected the limited partner course delivered by the Private Equity Institute at the Saïd Business School, University of Oxford, that led to the EVCA-awarded CIPEI. He is a Shimomura Fellow of the Development Bank of Japan and was a visiting researcher at Hitotsubashi University in Tokyo. Other career stations include intelligence officer in the German Air Force and CFO of Allianz Asia Pacific in Singapore. Mr. Meyer has published several books on investment strategies and risk management for real assets. He has authored *Private Equity Unchained* (2015, Palgrave MacMillan) and is the coauthor of *Beyond the J Curve* (2005, John Wiley & Sons), *J Curve Exposure* (2007, John Wiley & Sons), and *Mastering Illiquidity* (2011, John Wiley & Sons). Thomas Meyer and Pierre-Yves Mathonet are the primary authors of Chapters 7–9 and 11–13 (Private Equity and Venture Capital).

Putri Pascualy is a Partner and Managing Director at PAAMCO. She manages the firm's Long/Short Credit Portfolio and is the Portfolio Manager for custom portfolios for leading institutional investors. Ms. Pascualy leads the firm's investment efforts in corporate credit including high-yield bonds, bank loans, event-driven and opportunistic credit, distressed debt, and structured products. In addition to her research responsibilities, her expertise includes portfolio construction, structuring, and risk management of complex portfolios and investments throughout various market cycles. She graduated from UC Berkeley with a BA in Economics and an MBA from the Haas School of Business. Putri is also a frequent contributor to media outlets including the *Wall Street Journal*, Bloomberg and Bloomberg Television, *U.S. News and World Report*, *Barron's*, the *Financial Times*, and CNBC. She is the author of *Investing in Credit Hedge Funds: An In-Depth Guide to Building Your Portfolio and Profiting from the Credit Market* (2013, McGraw-Hill). She is the primary author of Chapter 29 (Hedge Funds: Credit Strategies).

Jason Scharfman is a Managing Partner of Corgentum Consulting, LLC. Corgentum is a specialty consulting firm that performs operational due diligence reviews and background investigations on fund managers of all types globally including hedge funds, private equity, and real estate funds. Mr. Scharfman is recognized as one of the leading experts in the field of operational due diligence and is the author of *Hedge Fund Governance: Evaluating Oversight, Independence, and Conflicts* (2014, Academic Press), *Private Equity Operational Due Diligence: Tools to Evaluate Liquidity, Valuation, and Documentation* (2012, John Wiley & Sons) *and Hedge Fund Operational Due Diligence: Understanding the Risks* (2008, John Wiley & Sons). Before founding Corgentum, he oversaw the operational due diligence function for a $6 billion alternative investment allocation group called Graystone Research at Morgan Stanley. Prior to joining Morgan Stanley, he held positions at Lazard Asset Management, SPARX Investments and Research, and Thomson Financial. Mr. Scharfman received a BS in Finance with an additional major in Japanese from Carnegie Mellon University, an MBA in Finance from Baruch College's Zicklin School of Business, and a JD from St. John's School of Law. He is the primary author of Chapters 10 and 33 (Private Equity and Hedge Fund Operational Due Diligence).

Ed Szado is an Assistant Professor of Finance at Providence College and the Director of Research at the Institute for Global Asset and Risk Management. Dr. Szado earned a PhD in Finance from the Isenberg School of Management, University of Massachusetts–Amherst, an MBA from Tulane University, and a BComm from McMaster University. He has taught at Boston University, Clark University, Providence College, and the University of Massachusetts–Amherst. He is a former options trader and has worked extensively on asset allocation and risk managed investment programs. He was a founding coeditor of the *Alternative Investment Analyst Review* (*AIAR*) and currently a member of the editorial board of *The Journal of Alternative Investments* (*JAI*). He is a CFA Charter Holder and has consulted to the Options Industry Council, the Chicago Board Options Exchange, the Chartered Alternative Investment Analyst Association, and the Commodity Futures Trading Commission. He is the primary author of Chapters 22–24 (Commodities).

PART 1

Asset Allocation and Institutional Investors

CHAPTER 1

Asset Allocation Processes and the Mean-Variance Model

This is the first of two chapters discussing asset allocation, with a focus on the decision-making process of asset allocators who consider portfolios consisting of traditional as well as alternative asset classes. This chapter describes the basic steps of the asset allocation process followed by a typical asset allocator. The objectives and constraints that apply to different types of asset owners are presented, and the important features of strategic and tactical asset allocation approaches are discussed. The chapter then explains the mean-variance approach, which is the best-known quantitative approach to allocation. Finally, some important limitations of the mean-variance approach are discussed.

1.1 IMPORTANCE OF ASSET ALLOCATION

Asset allocation refers both to the process followed by a portfolio manager to determine the distribution of an investor's assets to various asset classes and to the resulting portfolio weights. The allocation is determined to meet one or more objectives subject to a set of constraints set by the investor or dictated by the markets. An objective might be to maximize the expected value of a portfolio at a certain date subject to a set of constraints either established by the investor, such as a maximum level of return volatility or a maximum exposure to certain sectors, or dictated by the markets, such as no short selling of certain assets and a minimum investment level demanded by hedge fund managers.

While asset allocation refers to composition of an investor's portfolio in terms of different asset classes, we define **security selection** as the process through which holdings within each asset class are determined. For example, the asset allocation process may suggest that 20% of an investor's portfolio should be allocated to hedge funds, while security selection in this case is concerned with the hedge fund managers that are eventually selected for the investment purpose.

The importance of asset allocation versus security selection has been the subject of a long-running and controversial debate. The basic question is: Which of these two decisions has a larger impact on a portfolio's performance? As it turns out, the answer to this seemingly simple question is not that simple and, in some sense, it is impossible to provide.

First, we must specify whether the performance of a diversified or a concentrated portfolio is being measured. Clearly, the performance of a concentrated portfolio that consists of some allocation to cash and the rest to a single stock is mostly determined by the security selection decision. A significant portion of the characteristics of this portfolio's performance through time will depend on the choice of the single stock that constitutes the risky part of the portfolio. The choice of allocating a portion of the portfolio to cash will have some impact on the portfolio's performance, but it will be relatively small. In contrast, security selection is likely to have only a minor impact on the portfolio's performance if its equity portion consists of several thousand stocks that are listed around the world.

Second, we need to specify what is meant by portfolio performance. Is the impact of asset allocation on expected monthly return the sole criterion for evaluating the importance of asset allocation? How about higher moments of the return distribution or the beta of the portfolio with respect to some benchmark? As will be discussed, what is meant by performance will have an impact on the importance of asset allocation.

One of the most notable studies on the importance of asset allocation was published in 1986 by Brinson, Hood, and Beebower (BHB). The authors regressed the quarterly rates of return reported by a group of U.S. pension funds against passively managed benchmarks that were created using the weights proposed by the investment policy statements of the pension funds. The goal was to examine the relationship between the actual performance of the funds and the performance that would have been realized had the funds invested their capital in passively managed market indices according to the weights set forth in their investment policy statements. The average *r*-squared of these regressions exceeded 90%. Although BHB were clear in presenting their results, the rest of the investment community took the reported *r*-squared figure and made the blanket statement that more than 90% of the performance of these pension funds could be explained by the asset allocation decision described in the investment policy and that less than 10% of the performance could be explained by the active management decisions of the portfolio managers, such as security selection and tactical tilts. This would be the right conclusion if by performance one means the return *volatility* of the portfolio through time. However, this would be an incorrect conclusion if by performance one means the average return itself through time. In other words, BHB never claimed that 90% of the average return on diversified portfolios could be explained by the asset allocation decision.

As discussed in the CAIA Level I book, the *r*-squared of the regression tells how much of the variation in the dependent variable can be explained by variations in the independent or explanatory variables. In other words, the BHB study only confirmed that more than 90% of variability in the realized returns of fully diversified portfolios could be explained by the asset allocation decision. More important, it did not say anything about the impact of asset allocation on the average return on those pension funds. The study had a lot to say about the second moment of the funds' return distribution and very little about the first moment of their return distribution. Further, the sample included fully diversified portfolios and therefore could not consider the importance of security selection because the portfolio managers had already decided to fully diversify and not to hold concentrated positions in securities that they considered to be undervalued. In short, the study was not meant to answer some of the most important questions faced by asset allocators, but it did spur a large set of studies that have gradually provided answers to practitioners.

Three important questions that could be asked and answered regarding the importance of asset allocation for the performance of diversified portfolios are:

1. How much of the variability of returns across time is explained by the asset allocation framework set forth in the investment policy? That is, how many of a fund's ups and downs are explained by its policy benchmarks? The impact of asset allocation on time variation was studied in BHB. Since then, a number of studies have reexamined this question (Ibbotson and Kaplan 2000). These studies generally agree that a high degree (85% to 90%) of the time variation in diversified portfolios of traditional assets is explained by the overall asset allocation decisions of asset owners and portfolio managers. Therefore, if an asset allocator wants to evaluate the expected volatility of two diversified portfolios, then the asset allocation policies of the two funds will be very informative.
2. How much of the difference in the average returns among funds is explained by differences in the investment policy? That is, if the average returns of two diversified funds are compared, how much of the difference in relative performance can be explained by differences in asset allocation policies? The answer depends greatly on the sample, but most studies show that less than 50% of the difference in average returns can be explained by differences in asset allocation. Other factors—such as asset class timing, style within asset classes, security selection, and fees—explain the remaining differences. Therefore, if an asset allocator wants to evaluate the expected returns of two diversified funds, asset allocation policies of the two funds will be useful, but other factors should be taken into account.
3. What portion of the average return of a fund is explained by its asset allocation policy? In this case, we are considering the absolute performance of a fund. That is, suppose the realized average return on a fund is compared with the return on the fund if the manager had implemented the proposed asset allocation using passive benchmarks. How do these two performances compare? Does the manager outperform the passive implementation of the asset allocation policy? This appears to be the most relevant question, because it directly tests the active management of the portfolio. It turns out that this is the most difficult question to answer, and the available results are highly dependent on the sample and the period they cover. Most studies find that asset allocation has little explanatory power in predicting whether a manager will outperform or underperform the asset allocation return. In fact, available studies covering samples of mutual funds and pension funds conclude that 65% to 85% of them underperform the long-run asset allocation described in their investment policy statements or their passive benchmarks (Ibbotson and Kaplan 2000).[1]

Given the importance of asset allocation, the rest of this chapter focuses on the asset allocation process, the role of asset owners in determining the objectives and constraints of the process, and the difference between strategic and tactical asset allocation programs.

1.2 THE FIVE STEPS OF THE ASSET ALLOCATION PROCESS

This section describes the typical steps that must be taken to implement a systematic asset allocation program.[2] A systematic approach enables the asset allocator to design and implement an investment strategy for the sole benefit of the asset owners. Such an approach needs to focus on the objectives and the constraints that are relevant to the asset owner. We begin with a discussion of the first of the five steps in the asset allocation process: identifying the asset owners and their potential objectives and constraints. In most cases, assets are managed to fund potential liabilities. In some instances, these liabilities represent legal obligations of the asset owner, such as the assets of a defined benefit (DB) pension fund. In other cases, assets are not meant to fund legal obligations but to fund essential needs of the asset owners or their beneficiaries. For example, a foundation's assets are managed to fund its future philanthropic and grant-giving activities. The nature of these potential needs or liabilities is a major determinant of the objectives and constraints of each asset owner.

The second step involves developing an overall approach to asset allocation. A critical step is preparing the investment policy statement. The **investment policy statement** includes the asset allocator's understanding of the objectives and constraints of the asset owners, the menu of asset classes to be considered, whether active or passive approaches will be used, and how often and under what circumstances the allocation will be changed. Such changes arise because of fundamental changes in economic conditions or changes in the circumstances of the asset owner.

The third step is implementing the overall asset allocation policy described in the investment policy statement. This step will require applications of both quantitative and qualitative techniques to determine the weight of each asset class in the portfolio. Since allocations to alternative investments typically involve selection and allocation to managers (e.g., hedge fund and private equity managers), this step will need to have built-in flexibility, as extensive due diligence on managers must be completed, and thus planned allocations may turn out to be infeasible. For instance, the planned allocation may turn out to be less than the minimum investment level accepted by the manager who has emerged on top after the due diligence process.

The fourth step is allocating the capital according to the optimal weights determined in the previous step based on the due diligence and manager evaluation already conducted by the portfolio manager's team or outside consultants.

The final step is monitoring and evaluating the investments. Inevitably, the realized performance of the portfolio will turn out to be different than expected. This will happen because of unexpected changes in the market and because selected fund managers did not perform as expected. As previously stated, the investment policy statement should anticipate circumstances under which the allocation will be revised. This chapter focuses on the first four steps of the asset allocation process. The final step, which deals with benchmarking, due diligence, monitoring, and manager selection, was covered in CAIA Level I (benchmarking) and the rest of this book (due diligence, monitoring, and manager selection).

1.3 ASSET OWNERS

A systematic asset allocation process starts with the asset owners. Chapters 3 through 6 of this book provide detailed descriptions of major types of asset owners and their investment strategies. This section briefly describes major classes of asset owners. Although the list of asset owners will not be exhaustive, it should be sufficient to highlight the differences that exist among major types of asset owners and how their characteristics influence their asset allocation policies. The following sections discuss four categories of asset owners:

1. Endowments and foundations
2. Pension funds
3. Sovereign wealth funds
4. Family offices

1.3.1 Endowments and Foundations

Endowments and foundations serve different purposes but, from an investment policy point of view, share many characteristics. **Endowments** are funds established by not-for-profit organizations to raise funds through charitable contributions of supporters and use the resources to support activities of the sponsoring organization. For example, a university endowment receives charitable contributions from its supporters (e.g., alumni) and uses the income generated by the fund to support the normal operations of the university. Endowments could be small or large, but since they have long investment horizons and are lightly regulated, the full menu of assets is available to them. In fact, among institutional investors, endowments are pioneers in allocating to alternative assets.

Foundations are similar to endowments in the sense that funds are raised through charitable contributions of supporters. These funds are then used to fund grants and support other charitable work that falls within the foundation's mandate. Most foundations are long-term investors and are lightly regulated in terms of their investment activities. However, in order to enjoy certain tax treatments, they are required to distribute a minimum percentage of their assets each year. Foundations are able to invest in the full menu of assets, including alternative asset classes.

1.3.2 Pension Funds

Pension funds are set up to provide retirement benefits to a group of beneficiaries who typically belong to an organization, such as for-profit or not-for-profit businesses and government entities. The organization that sets up the pension fund is called the plan sponsor. There are four types of pension funds (Ang 2014):

1. NATIONAL PENSION FUNDS. **National pension funds** are run by national governments and are meant to provide basic retirement income to the citizens of a country. The U.S. Social Security program, South Korea's National Pension Service, and the Central Provident Fund of Singapore are examples of such funds. These types of funds may not operate that differently from sovereign wealth funds,

which are described later in this chapter and in Chapter 5 of this book. The investment allocation decisions of these large funds are controlled by national governments, which makes their management different from private pension funds. Given the size and long-term horizons of these funds, the menu of assets that are available for potential investments is large and includes various alternative assets.

2. Private defined benefit funds. **Private defined benefit funds** are set up to provide prespecified pension benefits to employees of a private business. The plan sponsor promises the employees of the private entity a predefined retirement income, which is based on a set of predetermined factors. Typically, these factors include the number of years an employee has worked for the firm, as well as his or her age and salary. The plan may include provisions for changes in retirement income, such as a cost-of-living adjustment or a portion of the retirement income to be paid to the employee's surviving spouse or young children. The plan sponsor directs the management of the fund's assets. While these funds may not match the size or the length of time horizon of national funds, they are still large long-term investors, and therefore the full menu of asset classes, including alternative assets, are available to them.
3. Private defined contribution funds. **Private defined contribution funds** are set up to receive contributions made by the plan sponsor into the fund. The pension plan specifies the contributions that the plan sponsor is expected to make while the firm employs the beneficiary. The contributions are deposited into accounts that are tied to each beneficiary, and upon retirement, the employee receives the accumulated value of the account. The employee and the plan sponsor jointly manage the fund's assets, in that the sponsor decides on the menu of asset classes available, and the employee decides the asset allocation. The menu of asset classes available to these funds is smaller than both national funds and defined benefit funds. Lumpiness of alternative investments, lack of liquidity, and government regulations typically prevent these funds from investing in a full range of alternative asset classes. Historically, real estate is one alternative asset class that has been available to these funds. In recent years, liquid alternatives have slowly become available as well.
4. Individually managed accounts. **Individually managed accounts** are no different from private savings plans, in which the asset allocation is directed entirely by the employee. Since the funds enjoy tax advantages, they are not free from regulations, and therefore the list of asset classes available to the beneficiary will be limited. In particular, privately placed alternative investments are not normally available to these funds.

1.3.3 Sovereign Wealth Funds

Sovereign wealth funds (SWFs) are funds set by national governments as a way to save and build on a portion of the country's current income for use by future generations of its citizens. SWFs are similar to national pension funds in the sense that they are owned and managed by national governments, but the goal is not to provide retirement income to the citizens of the country.

SWFs have become major players in global financial markets because of their sheer size and their long-term investment horizons. Most SWFs invest a portion of their assets in foreign assets. SWFs are relatively new, and their growth, especially in emerging economies, has been tied to the rise in prices of natural resources such as oil, copper, and gold. In some cases, SWFs are funded through the foreign currency reserves earned by countries that enjoy a significant trade surplus, such a China.

SWFs are large and have very long horizons; therefore the full menu of assets should be available to them. However, because national governments manage them, they may not invest in all available asset classes.

1.3.4 Family Offices

Family offices refer to organizations dedicated to the management of a pool of capital owned by a wealthy individual or group of individuals. In effect, it is a private wealth advisory firm established by an ultra-high-net-worth individual or family.

The source of income for a family office can be as varied as the underlying family that it serves. In some cases, the capital is spun off from an operating company, while in other cases, it might be funded with what is known as legacy wealth, which refers to a second or third generation of family members that have inherited their wealth from a prior source of capital generation. The financial resources of a family office can be used for a variety of purposes, from maintaining the family's current standard of living to providing benefits for many future generations to distributing all or a portion of it through philanthropic activities in the current generation. Family offices tend to have relatively long time horizons and are typically large enough to invest in a full menu of assets, including alternative asset classes.

1.4 OBJECTIVES AND CONSTRAINTS

As already discussed, different asset owners have their own particular objectives in managing their assets and face various constraints, which could be internal or external. An **objective** is a preference that distinguishes an optimal solution from a sub-optimal solution. A **constraint** is a condition that any solution must meet. Internal constraints are imposed by the asset owner and may be a function of the owner's time horizon, liquidity needs, and desire to avoid certain sectors. External constraints result from market conditions and regulations. For instance, an asset owner may be prohibited from investing in certain asset classes, or fees and due diligence costs may prevent the owner from considering all available asset classes. The next sections describe the issues that must be considered while attempting to develop a systematic understanding of asset owners' objectives and constraints.

1.5 INVESTMENT POLICY OBJECTIVES

Asset owners' objectives must be expressed in terms of consistent risk-adjusted performance values. In other words, it is safe to assume that asset owners would prefer to earn a high rate of return on their assets. However, higher rates of return are associated with higher levels of risk. Therefore, asset owners should present their objectives in terms of combinations of risks and returns that are consistent with market conditions and their level of risk tolerance. For instance, the objective of earning 30%

per year on a portfolio that has 8% annual volatility is not consistent with market conditions. Such a high return would require a much higher level of volatility. Also, if the asset owner states that her objective is to earn 25% per year with no reference to the level of risk that she is willing to assume, then it could lead the portfolio manager to create a risky portfolio that is entirely inconsistent with her risk tolerance. Therefore, asset owners and portfolio managers need to communicate in a clear language regarding return objectives and risk levels that are acceptable to the asset owner and are consistent with current market conditions.

1.5.1 Evaluating Objectives with Expected Return and Standard Deviations

Consider the following two investment choices available to an asset owner:

- Investment A will increase by 10% or decrease by 8% over the next year, with equal probabilities.
- Investment B will increase by 12% or decrease by 10% over the next year, with equal probabilities.
- The expected return on both investments is 1% (found as the probability weighted average of their potential returns); however, their volatilities will be different (see Equation 4.9 of CAIA Level I).
- Investment A: Standard deviation = $\sqrt{0.5\times(0.10-0.01)^2+0.5\times(-0.08-0.01)^2}$ = 9%
- Investment B: Standard deviation = $\sqrt{0.5\times(0.12-0.01)^2+0.5\times(-0.10-0.01)^2}$ = 11%

If an asset owner expresses a preference for investment A over investment B, then we can claim that the asset owner is risk averse. Although it is rather obvious to see why a risk-averse asset owner would prefer A to B, it will not be easy to determine whether a risk-averse investor would prefer C to D from the following example:

- Investment C will increase by 10% or decrease by 8% over the next year, with equal probabilities.
- Investment D will increase by 12% or decrease by 9% over the next year, with equal probabilities.

In this case, compared to investment C, investment D has a higher expected return (1.5% to 1%) and a higher standard deviation (10.5% to 9%). Depending on their aversion to risk, some asset owners may prefer C to D, and others, D to C.

1.5.2 Evaluating Risk and Return with Utility

Different asset owners will have their own preferences regarding the trade-off between risk and return. Economists have developed a number of tools for expressing such preferences. Expected utility is the most common approach to specifying the preferences of an asset owner for risk and return. While a utility function is typically used to express preferences of individuals, there is nothing in the theory or application that would prevent us from applying this to institutional investors as well. Therefore, in the context of investments, we define **utility** as a measurement of the satisfaction that an individual receives from investment wealth or return. **Expected utility** is the probability weighted average value of utility over all possible outcomes. Finally, in the context of investments, a **utility function** is the relationship that converts an investment's financial outcome into the investor's level of utility.

Suppose the initial capital available for an investment is W and that the utility derived from W is $U(W)$. Thus, the expected utilities associated with investments A and B can be expressed as follows:

$$E[U(W_A)] = 0.5 \times U(W \times 1.10) + 0.5 \times U(W \times 0.92) \tag{1.1}$$

$$E[U(W_B)] = 0.5 \times U(W \times 1.12) + 0.5 \times U(W \times 0.90) \tag{1.2}$$

The function $U(\bullet)$ is the utility function. The asset owner would prefer investment A to investment B if $E[U(W_A)] > E[U(W_B)]$.

Suppose the utility function can be represented by the log function, and assume that the initial investment is \$100. Then:

$$E[U(W_A)] = 0.5 \times \ln(100 \times 1.10) + 0.5 \times \ln(100 \times 0.92) = 4.611 \tag{1.3}$$

$$E[U(W_B)] = 0.5 \times \ln(100 \times 1.12) + 0.5 \times \ln(100 \times 0.90) = 4.609 \tag{1.4}$$

In this case the asset owner would prefer investment A to investment B because it has higher expected utility. Applying the same function to investments C and D, it can be seen that $E[U(W_C)] = 4.611$ and $E[U(W_D)] = 4.615$. In this case, the asset owner would prefer investment D to investment C.

APPLICATION 1.5.2

Suppose that an investor's utility is the following function of wealth (W):

$$U(W) = \sqrt{W}$$

Find the current and expected utility of the investor if the investor currently has \$100 and is considering whether to speculate all the money in an investment with a 60% chance of earning 21% and a 40% chance of losing 19%. Should the investor take the speculation rather than hold the cash?

The current utility of holding the cash is 10, which can be found as $\sqrt{100}$. The expected utility of taking the speculation is found as:

$$E[U(W)] = (0.60 \times \sqrt{121}) + (0.40 \times \sqrt{81}) = 10.2$$

Because the investor's expected utility of holding the cash is only 10, the investor would prefer to take the speculation, which has an expected utility of 10.2.

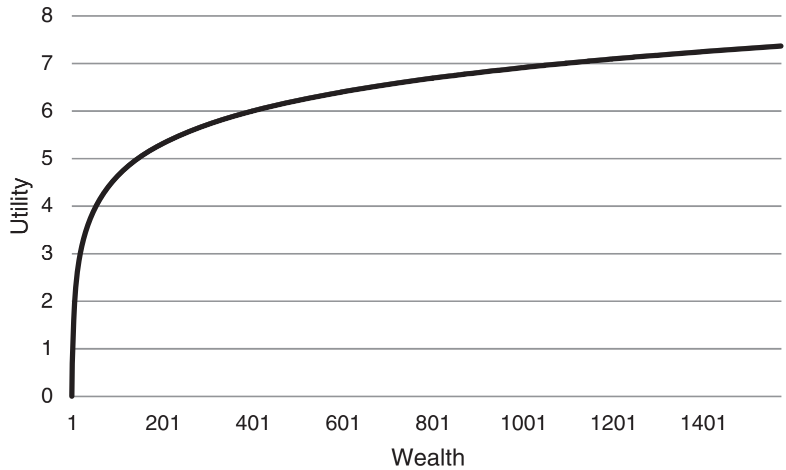


EXHIBIT 1.1 Logarithmic Utility Function

1.5.3 Risk Aversion and the Shape of the Utility Function

We are now prepared to introduce a more precise definition of risk aversion. An investor is said to be **risk averse** if his utility function is concave, which in turn means that the investor requires higher expected return to bear risk. Exhibit 1.1 displays the log function for various values of wealth. We can see that the level of utility increases but at a decreasing rate.

Alternatively, a risk-averse investor avoids taking risks with zero expected payoffs. That is, for risk-averse investors, $E[U(W)] > E[U(W + \tilde{e})]$, where $\tilde{e}$ is a zero mean random error that is independent from W.

1.5.4 Expressing Utility Functions in Terms of Expected Return and Variance

The principle of selecting investment strategies and allocations to maximize expected utility provides a very flexible way of representing the asset owner's preferences for risk and return. The representation of expected utility can be made more operational by presenting it in terms of the parameters of the probability distribution functions of investment choices. The most common form among institutional investors is to present the expected utility of an investment in terms of the mean and variance of the investment returns. That is,

$$E[U(W)] = \mu - \frac{\lambda}{2} \times \sigma^2 \tag{1.5}$$

Here, μ is the expected rate of return on the investment, σ^2 is the variance of the rate of return, and λ is a constant that represents the asset owner's degree of risk aversion. It can be seen that the higher the value of λ, the higher the negative effect of variance on the expected value. For example, if λ is equal to zero, then the investor is said to be risk neutral and the investment is evaluated only on the basis of its expected return. A negative value of λ would indicate that the investor is a risk seeker and actually prefers more risk to less risk.

The **degree of risk aversion** indicates the trade-off between risk and return for a particular investor and is often indicated by a particular parameter within a utility function, such as λ in Equation 1.5. The fact that the degree of risk aversion is divided by 2 will make its interpretation much easier. It turns out that if Equation 1.5 is used to select an optimal portfolio for an investor, then the ratio of the expected rate of return on the optimal portfolio in excess of the riskless rate divided by the portfolio's variance will be equal to the degree of risk aversion.

Example: Suppose $\lambda = 5$. Calculate the expected utility of investments C and D.

$$\mu_C - \frac{\lambda}{2} \times \sigma_C^2 = 1.0\% - (5/2) \times (9.0\%)^2 = -0.01025$$

$$\mu_D - \frac{\lambda}{2} \times \sigma_D^2 = 1.5\% - (5/2) \times (10.5\%)^2 = -0.01256$$

In this case, the expected utility of investment C is higher than that of investment D; therefore, it is the preferred choice. It can be verified that if $\lambda = 1$, then the expected utility of investments C and D will be 0.0059 and 0.00949, respectively, meaning that D will be preferred to C.

APPLICATION 1.5.4

Suppose that an investor's expected utility, $E[U(W)]$, from an investment can be expressed as:

$$E[U(W)] = \mu - \frac{\lambda}{2} \times \sigma^2$$

where W is wealth, μ is the expected rate of return on the investment, σ^2 is the variance of the rate of return, and λ is a constant that represents the asset owner's degree of risk aversion.

Use the expected utility of an investor with $\lambda = 0.8$ to determine which of the following investments is more attractive:

Investment A: $\mu = 0.10$ and $\sigma^2 = 0.04$

Investment B: $\mu = 0.13$ and $\sigma^2 = 0.09$

The expected utility of A and B are found as:

Investment A: $E[U(W)] = 0.10 - \frac{0.8}{2} \times 0.04 = 0.084$

Investment B: $E[U(W)] = 0.13 - \frac{0.8}{2} \times 0.09 = 0.094$

Because the investor's expected utility of holding B is higher, investment B is more attractive.

EXHIBIT 1.2 Properties of Two Hedge Fund Indices

Index	Annualized Mean	Annualized Std. Dev.	Skewness
HFRI Fund of Fund Defensive	7.09%	5.70%	0.24
HFRI Fund Weighted Composite	9.77%	6.76%	–0.61

Source: HFR and authors' calculations.

1.5.5 Expressing Utility Functions with Higher Moments

When the expected utility is presented as in Equation 1.5, we are assuming that risk can be measured using variance or standard deviation of returns. This assumption is reasonable if investment returns are approximately normal. While the normal distribution might be a reasonable approximation to returns for equities, empirical evidence suggests that most alternative investments have return distributions that significantly depart from the normal distribution. In addition, return distributions from structured products tend to deviate from normality in significant ways. In these cases, Equation 1.5 will not be appropriate for evaluating investment choices that display significant skewness or excess kurtosis. It turns out that Equation 1.5 can be expanded to accommodate asset owners' preferences for higher moments (i.e., skewness and kurtosis) of return distributions. For example, one may present expected utility in the following form:

$$E[U(W)] = \mu - \frac{\lambda_1}{2} \times \sigma^2 + \lambda_2 \times S - \lambda_3 \times K \tag{1.6}$$

Here, S is the skewness of the portfolio value; K is the kurtosis of the portfolio; and $\lambda_1, \lambda_2,$ and λ_3 represent preferences for variance, skewness, and kurtosis, respectively. It is typically assumed that most investors dislike variance ($\lambda_1 > 0$), like positive skewness ($\lambda_2 > 0$), and dislike kurtosis ($\lambda_3 > 0$). Note that the signs of coefficients change.

Example: Consider the information about two hedge fund indices in Exhibit 1.2.

If we set $\lambda_1 = 10$ and ignore higher moments, the investor would select the HFRI Fund Weighted Composite as the better investment, as it would have the higher expected utility (0.075 to 0.055). However, if we expand the objective function to include preference for positive skewness and set $\lambda_2 = 1$, then the investor would select the HFRI Fund of Fund Defensive as the better choice, because it would have a higher expected utility (0.29 to –0.54).

1.5.6 Expressing Utility Functions with Value at Risk

The preceding representation of preferences in terms of moments of the return distribution is the most common approach to modeling preferences involving uncertain choices. It is theoretically sound as well. However, the investment industry has developed a number of other measures of risk, most of which are not immediately comparable to the approach just presented. For instance, in the CAIA Level I book, we learned about value at risk (VaR) as a measure of downside risk. Is it possible to use this framework to model preferences in terms of VaR? It turns out that in a rather ad hoc way, one can use the preceding approach to model preferences on risk and return when risk is measured by VaR. That is, we can rank investment choices by calculating the following value:

$$E[U(W)] = \mu - \frac{\lambda}{2} \times \text{VaR}_\alpha \tag{1.7}$$

Here, λ can be interpreted as the degree of risk aversion toward VaR, and VaR_α is the value at risk of the portfolio with a confidence level of α.

We can further generalize Equation 1.7 and replace VaR with other measures of risk. For instance, one could use risk statistics, such as lower partial moments, beta with respect to a benchmark, or the expected maximum drawdown.

1.5.7 Using Risk Aversion to Manage a Defined Benefit Pension Fund

To complete our discussion of objectives, we now consider an application of the previous framework to present the objectives of a defined benefit (DB) pension fund. The following information is available:

- Current value of the fund: €V billion
- Number of asset classes considered: N
- Return on asset class i: R_i
- Weight of asset class i in the portfolio: w_i
- Return on the portfolio: $R_p = \sum_{i=1}^{N} w_i R_i$

Assuming that the preferences of the DB fund can be expressed as in Equation 1.5, the portfolio manager will select the weights, w_i, such that the expected utility is maximized. That is, Equation 1.8 expresses the objective function that is maximized by choosing the values of w_i. Of course, the portfolio manager must ensure that the weights will add up to one and some or all of the weights will need to be positive.

$$E[U(W)] = V \times E[R_p] - \frac{\lambda}{2} \times \text{Var}[V \times R_p] \tag{1.8}$$

1.5.8 Finding Investor Risk Aversion from the Asset Allocation Decision

As mentioned previously, the value of the risk aversion has an intuitive interpretation. The expected excess rate of return on the optimal portfolio ($E[R_P] - R_f$) divided by its variance, σ_P^2, is equal to the degree of risk aversion, λ:

$$\lambda = \frac{E[R_P] - R_f}{\sigma_P^2} \tag{1.9}$$

The value of the parameter of risk aversion, λ, is chosen in close consultation with the plan sponsor. There are qualitative methods that can help the portfolio manager

EXHIBIT 1.3 Hypothetical Risk Returns for Two Portfolios

Portfolio	Annualized Mean	Annualized Std. Dev.
Aggressive	15%	16%
Moderate	9%	8%

select the appropriate value of the risk aversion. The portfolio manager may select a range of values for the parameters and present asset owners with resulting portfolios so that they can see how their level of risk aversion affects the risk-return characteristics of the portfolio under current market conditions.

Example: Consider the information for two well-diversified portfolios shown in Exhibit 1.3. The riskless rate is 2% per year.

Assuming that these are optimal portfolios for two asset owners, what are their degrees of risk aversion?

We know from Equation 1.9 that the expected excess return on each portfolio divided by its variance will be equal to the degree of the risk aversion of the investor who finds that portfolio optimal.

Aggressive investor: $(15\% - 2\%)/(16\%^2) = 5.1$

Moderate investor: $(9\% - 2\%)/(8\%^2) = 10.9$

As expected, the aggressive portfolio represents the optimal portfolio for a more risk-tolerant investor, while the moderate portfolio represents the optimal portfolio for a more risk-averse investor.

APPLICATION 1.5.8

Suppose that an investor's optimal portfolio has an expected return of 10%, which is 8% higher than the riskless rate. If the variance of the portfolio is 0.04, what is the investor's degree of risk aversion, λ?

Using Equation 1.9, λ can be expressed as:

$$\lambda = \frac{E[R_P] - R_f}{\sigma_P^2} = \frac{0.08}{0.04} = 2$$

1.5.9 Managing Assets with Risk Aversion and Growing Liabilities

As mentioned earlier in the chapter, most asset owners are concerned with funding future obligations using the income generated by the assets. In the previous example, the DB plan has liabilities that will need to be met using the fund's assets. Suppose the current value of these liabilities is L euros. Further, suppose the rate of growth in liabilities is given by G, which could be random. In this case, the objective function of Equation 1.8 can be restated as:

$$E[U(W)] = V \times E[R_p] - \frac{\lambda}{2} \times \mathrm{Var}[V \times R_p - L \times G] \tag{1.10}$$

In this case, the DB plan wishes to maximize the expected rate of return on the fund's assets, subject to its aversion toward deviations between the return on the fund and the growth in the fund's liabilities. In other words, the risk of the portfolio is measured relative to the growth in liabilities. Later in this chapter, we will demonstrate how this problem can be solved.

One final comment about evaluating investment choices: Although the framework outlined here is a flexible and relatively sound way of modeling preferences for risk and return, the presentation considered only one-period investments and decisions. Economists have developed methods for extending the framework to more than one period, where the investor has to withdraw some income from the portfolio. These problems are extremely complex and beyond the scope of this book. However, in many cases, the solutions that are based on the single-period approach provide a reasonable approximation of the solutions obtained under approaches that are more complex.

1.6 INVESTMENT POLICY CONSTRAINTS

The previous section introduced the expected utility approach as a simple and yet flexible approach to modeling risk-return objectives of asset owners. This section discusses the typical set of constraints that must be taken into account when trying to select the investment strategy that maximizes the expected utility of the asset owner.

1.6.1 Investment Policy Internal Constraints

Internal constraints refer to those constraints that are imposed by the asset owner as a result of its specific needs and circumstances. Some of these internal constraints can be incorporated into the objective function previously discussed. For example, we noted how the constraint that allocations with positive skewness are preferred could be incorporated into the model. However, there are other types of constraints that may be expressed separately. Some examples of these internal constraints are:

- Liquidity. The asset owner may have certain liquidity needs that must be explicitly recognized. For example, a foundation may be anticipating a large outlay in the next few months and therefore would want to have enough liquid assets to cover those outflows. This will require the portfolio manager to impose a minimum investment requirement for cash and other liquid assets. Even if there are no anticipated liquidity events where cash outlays will be needed, the asset owner may require maintaining a certain level of liquidity by imposing minimum investment requirements for cash and cash-equivalent investments, and maximum investment levels for such illiquid assets as private equity and infrastructure.
- Time horizon. The asset owner's investment horizon can affect liquidity needs. In addition, it is often argued that investors with a short time horizon should take

less risk in their asset allocation decisions, as there is not enough time to recover from a large drawdown. This impact of time horizon can be taken care of by changing the degree of risk aversion or by imposing a maximum limit on allocations to risky assets. Time horizon may impact asset allocation in other ways as well. For instance, certain asset classes are known to display mean reversion in the long run (e.g., commodities). As a result, an investor with a short time horizon may impose a maximum limit on the allocation to commodities, as there will not be enough time to enjoy the benefits of potential mean reversion.

- Sector and country limits. For a variety of reasons, an asset owner may wish to impose constraints on allocations to certain countries or sectors of the global economy. For instance, national pension plans may be prohibited from investing in certain countries, or a university endowment may have been instructed by its trustees to avoid investments in certain industries.

Asset owners may have unique needs and constraints that have to be accommodated by the portfolio manager. However, it is instructive to present asset owners with alternative allocations in which those constraints are relaxed. This will help asset owners understand the potential costs associated with those constraints.

1.6.2 Investment Policy and the Two Major Types of External Constraints

External constraints refer to constraints that are driven by factors that are not directly under the control of the investor. These constraints are mostly driven by regulations and the tax status of the asset owner.

- Tax status. Most institutional investors are tax exempt, and therefore allocation to tax-exempt instruments are not warranted. Because these investments offer low returns, the optimization technique selected to execute the investment strategy should automatically exclude those assets. In contrast, family offices and high-net-worth investors are not tax exempt, and therefore the impact of taxes must be taken into account. For example, constraints can be imposed to sell asset classes that have suffered losses to offset realized gains from those that have increased in value.
- Regulations. Some institutional investors, such as public and private pension funds, are subject to rules and regulations regarding their investment strategies. In the United States, the Employee Retirement Income Security Act (ERISA) represents a set of regulations that affect the management of private pension funds. In the United Kingdom, the rules and regulations set forth by the Financial Services Authority impact pension funds. In these and many other countries, regulations impose limits on the concentration of allocations in certain asset classes.

1.7 PREPARING AN INVESTMENT POLICY STATEMENT

The next step in the process is to develop the overall framework of the asset allocation by preparing an investment policy statement (IPS).[3]

1.7.1 Seven Common Components of an Investment Policy Statement

The policy may include a recommended strategic allocation as well. The following is an outline of a typical IPS based on seven common components.

1. Background. A typical IPS begins with the background of the asset owner and its mission. It reminds all parties who the beneficiaries of the assets are.
2. Objective. The overall goals of the asset owner are described. For instance, the IPS of a foundation may state that the broad objectives are to (1) maintain the purchasing power of the current assets and all future contributions, (2) achieve returns within reasonable and prudent levels of risk, and (3) maintain an appropriate asset allocation based on a total return policy that is compatible with a flexible spending policy while still having the potential to produce positive real returns. The IPS may also provide additional details about the level of risk tolerance, the investment horizon, and the level of expected return that is needed to meet certain liabilities.
3. Asset classes. This segment will include a list of asset classes that the portfolio manager is allowed to consider for allocation. It may provide additional information about how each asset class will be accessed. For instance, the asset owner may decide to use a passive approach to allocations to traditional asset classes and then use active managers for alternative asset classes.
4. Governance. The organizational structure of the fund is described here. The responsibilities of various parties who are involved in the investment process (e.g., the portfolio manager, investment committee, administrator, and custodian) are carefully explained.
5. Manager selection. This section describes the basic framework that the asset owner will follow in selecting outside managers. For example, it may state that all hedge fund managers will need to have three years of experience with at least $100 million in assets under management.
6. Reporting and monitoring. The IPS describes the reporting requirements for the portfolio manager (e.g., frequency, type of reports, and disclosures).
7. Strategic asset allocation. The IPS may include the long-run allocation of the fund during normal periods. The statement may include upper and lower limits for each asset class as well. Further details about strategic asset allocation are discussed in the next section.

1.7.2 Strategic Asset Allocation: Risk and Return

The central focus of strategic asset allocation (SAA) is to create a portfolio allocation that will provide the asset owner with the optimal balance between risk and return over a long-term investment horizon. The SAA not only represents the long-run normal allocation of the investors' assets but also serves as the basis for creating a benchmark that will be used to measure the actual performance of the portfolio. The SAA also serves as the starting point of the tactical asset allocation process, which will adjust the SAA based on short-term market forecasts.[4] (Tactical asset allocation will be discussed in the next chapter.)

SAA is based on long-term risk-return relationships that have been observed in the past and that, based on economic and financial reasoning, are expected to persist under normal economic conditions into the future. While historical risk-return relationships are used as the starting point of generating the inputs needed to create the optimal long-run allocation, these historical relationships should be adjusted to reflect fundamental and potentially long-lasting economic changes that are currently taking place. For example, although long-term historical returns to investment-grade corporate bonds were once high, the prevailing yields on those instruments would indicate that the long-run return from this asset class should be adjusted down.

In developing long-term risk-return relationships for major asset classes, it is important to begin with fundamental factors affecting the economy. Macroeconomic performance of the global economy is the driving force behind the performance of various asset classes. The expected return on all asset classes can be expressed as the sum of three components:

$$\begin{aligned}\text{Asset Class Return} = {} & \text{Short-Term Real Riskless Rate} \\ & + \text{Expected Inflation} + \text{Risk Premium}\end{aligned} \tag{1.11}$$

The real short-term riskless rate of interest is believed to be relatively stable and lower than the real growth rate in the economy.[5] Typically, there is a lower bound of zero for this rate. Therefore, if the global economy is expected to grow at 3% per year going forward, the short-term real riskless rate is expected to be somewhere between zero and 1%. In turn, population growth and increases in productivity are known to be the major drivers of economic growth. Long-term expected inflation is far less stable, as it depends on central banks' policies as well as long-term economic growth. Historically, it was believed that long-term expected inflation would depend on the growth rate in the supply of money relative to the real growth rate in the economy. For instance, it was believed that long-term inflation would be around 5% if the money supply were to grow at 8% in an economy that is growing at 3%. However, this long-term relationship has been challenged by empirical observations following the 2008–9 financial crisis.

Once long-term estimates of the short-term real riskless rate and expected inflation have been obtained, the next step involves the estimation of the long-term risk premium of each asset class. At this stage, one may assume that historical risk premiums would prevail going forward. This would be particularly appropriate if we believe that historical estimates of volatilities, correlations, and risk exposures of various asset classes are likely to persist into the future. For instance, if the long-term historical risk premium on small-cap equities has been 5%, then, assuming 2% expected inflation and a 1% short-term real riskless rate, one could assume an 8% expected long-term return from this asset class.

For several reasons, long-term returns from alternative asset classes could be more difficult to estimate. First, while alternative assets such as real estate and commodities have a long history, some of the more modern alternative asset classes (e.g., hedge funds or private equity) do not have a long-enough history to obtain accurate estimates of their risk exposures and risk premiums. Second, to the degree that alpha was a major source of return for alternative asset classes in the past, the same level of alpha may not be available going forward if there is increased allocation to this asset class by investors. That is, the supply of alpha is limited, and increased competition is bound to reduce it. Third, the alternative investment industry has shown to be quite

EXHIBIT 1.4 Hypothetical Strategic Asset Allocation for an Endowment

	Minimum Allocation	Strategic Allocation	Maximum Allocation
Cash and Short-Term Treasuries	2%	5%	10%
Long-Term Investment-Grade Bonds	10%	10%	20%
High-Yield Bonds	0%	5%	7%
Large-Cap Equities	5%	15%	30%
Small-Cap Equities	0%	10%	15%
Emerging and Frontier Equities	0%	10%	15%
Commodities	0%	5%	10%
Real Estate and Other Real Assets	5%	15%	20%
Private Equity	5%	15%	20%
Hedge Funds	5%	10%	20%
		100%	

innovative and adaptive in response to changing economic conditions. Therefore, we should expect to see new classes of alternative assets going forward, with their potential place in investors' strategic asset allocations unknowable at this point.

1.7.3 Developing a Strategic Asset Allocation

Given the risk-return preference of an asset owner and estimates of expected long-term returns from various asset classes, the portfolio manager and the asset owner can develop an SAA. Exhibit 1.4 displays a hypothetical SAA for a U.S. endowment.

A typical IPS contains a strategic asset allocation and describes the circumstances under which the strategic asset allocation could change; for example, due to fundamental changes in the global economy or changes in the circumstances of the asset owner, the SAA could be revised.

1.7.4 A Tactical Asset Allocation Strategy

Related to SAA is tactical asset allocation (TAA), which is a dynamic asset allocation strategy that actively adjusts a portfolio's SAA based on short- to medium-term market forecasts. TAA's objective is to systematically exploit temporary market inefficiencies and divergences in market values of assets from their fundamental values. Long-term performance of a broadly diversified portfolio is driven mostly by its SAA over time. TAA can add value if designed based on rigorous economic analysis of financial data so it can overcome the headwinds created by the costs associated with portfolio turnover and the fact that global financial markets are generally efficient. The next chapter will provide further details about TAA and the more recent developments based on factor allocation and economic regime-driven investment strategies.

1.8 IMPLEMENTATION

After the completion of the IPS, the next step is its implementation. A variety of quantitative and qualitative portfolio construction approaches are available for this stage. We will focus our attention on the mean-variance approach, as it is the best-known approach, and most of the subsequent developments in this area have attempted to improve on its shortcomings. Some of these approaches are discussed in the next chapter.

Earlier, this chapter discussed how the general expected utility approach could be used to represent preferences in terms of moments of a portfolio's return distribution. In particular, we noted that optimal portfolios could be constructed by selecting the weights such that the following function is maximized:

$$\mu - \frac{\lambda}{2} \times \sigma^2 \tag{1.12}$$

where μ is the expected return on the portfolio, λ is a parameter that represents the risk-aversion of the asset owner, and σ^2 is the variance of the portfolio's return. The next section provides a more detailed description of this portfolio construction technique and examines the solution under some specific conditions. Later sections will discuss some of the problems associated with this portfolio optimization technique and offer some of the solutions that have been proposed by academic and industry researchers.

1.8.1 Mean-Variance Optimization

The portfolio construction problem discussed in this section is the simplest form of mean-variance optimization. The universe of risky investments available to the portfolio manager consists of N asset classes. The single-period total rate of return on the risky asset i is denoted by R_i, for $i = 1, \dots, N$. We assume that asset zero is riskless, and its rate of return is given by R_0. The weight of asset i in the portfolio is given by w_i. Therefore, the rate of return on a portfolio of the $N + 1$ risky and riskless asset can be expressed as:

$$R_p = w_0 R_0 + w_1 R_1 + \cdots + w_N R_N \tag{1.13}$$

$$w_0 + w_1 + \cdots + w_N = 1 \tag{1.14}$$

For now, we do not impose any short-sale restriction, and therefore the weights could assume negative values.

From Equation 1.14, we can see that $w_0 = 1 - \sum_{i=1}^{N} w_i$. If this is substituted in Equation 1.13 and terms are collected, the rate of return on the portfolio can be expressed as:

$$R_p = w_1(R_1 - R_0) + \cdots + w_N(R_N - R_0) + R_0 \tag{1.15}$$

The advantage of writing the portfolio's rate of return in this form is that we no longer need to be concerned that the weights appearing in Equation 1.15 will add up to one. Once the weights of the risky assets are determined, the weight of the riskless asset will be such that all the weights would add up to one.

Next, we need to consider the risk of this portfolio. Suppose the covariance between asset i and asset j is given by σ_{ij}. Using this, the variance-covariance of the N risky assets is given by:

$$\sum = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \sigma_{ij} & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{bmatrix} \tag{1.16}$$

The portfolio problem can be written in this form, where the weights are selected to maximize the objective function:

$$\max_{w_1,\ldots w_N} E\left[\sum_{i=1}^{N} w_i(R_i - R_0) + R_0\right] - \frac{\lambda}{2} \times \mathrm{Var}\left[\sum_{i=1}^{N} w_i(R_i - R_0) + R_0\right] \tag{1.17}$$

This turns out to have a simple and well-known solution:

$$\begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \frac{1}{\lambda} \sum^{-1} \times \begin{bmatrix} E[R_1 - R_0] \\ \vdots \\ E[R_N - R_0] \end{bmatrix} \tag{1.18}$$

The solution requires one to obtain an estimate of the variance-covariance matrix of returns on risky assets. Then the inverse of this matrix will be multiplied into a vector of expected excess returns on the N risky assets. It is instructive to notice the role of the degree of risk aversion. As the level of risk aversion (λ) increases, the portfolio weights of risky assets decline. In addition, those assets with large expected excess returns tend to have the largest weights in the portfolio.

1.8.2 Mean-Variance Optimization with a Risky and Riskless Asset

To gain a better understanding of the solution, consider the case of only one risky asset and a riskless asset. In this case, the optimal weight of the risky asset using Equation 1.18 will be:

$$w = \frac{1}{\lambda} \frac{E[R - R_0]}{\sigma^2} \tag{1.19}$$

The optimal weight of the risky asset is proportional to its expected excess rate of return, $E[R - R_0]$, divided by its variance, σ^2. Again, the higher the degree of risk aversion, the lower the weight of the risky asset.

For example, with an excess return of 10%, a degree of risk aversion (λ) of 3, and a variance of 0.05, the optimal portfolio weight is 0.67. This is found as (1/3) × (0.10/0.05). Note that Equation 1.19 may be used to solve for any of the variables, given the values of the remaining variables.

APPLICATION 1.8.2

Consider the case of mean-variance optimization with one risky asset and a riskless asset. Suppose the expected rate of return on the risky asset is 9% per year. The annual standard deviation of the index is estimated to be 13% per year. If the riskless rate is 1%, what is the optimal investment in the risky asset for an investor with a risk-aversion degree of 10?

The solution is:

$$\begin{aligned} w &= \frac{1}{10} \times \frac{0.09 - 0.01}{0.13^2} = 47.3\% \\ w_0 &= 1 - 47.3\% = 52.7\% \end{aligned}$$

That is, this investor will invest 47.3% in the risky asset and 52.7% in the riskless asset. By varying the degree of risk aversion, we can obtain the full set of optimal portfolios.

1.8.3 Mean-Variance Optimization with Growing Liabilities

Equation 1.10 displayed the formulation of the problem when the asset owner is concerned with the tracking error between the value of the assets and the value of the liabilities. Similar to Equation 1.18, a general solution for that problem can be obtained as well. Here we present a simple version of it when there is only one risky asset. The covariance between the rate of growth in the liabilities and the growth in assets is denoted by δ, and L is the value of liabilities relative to the size of assets:

$$w = \frac{1}{\lambda} \frac{E[R - R_0]}{\sigma^2} + L\frac{\delta}{\sigma^2} \tag{1.20}$$

It can be seen that if the risky asset is positively correlated with the growth in liabilities (i.e., $\delta > 0$), then the fund will hold more of that risky asset. The reason is that the risky asset will help reduce the risk associated with growth in liabilities. For instance, if the liabilities behaved like bonds, then the fund would invest more in fixed-income instruments, as they would reduce the risk of the fund.

Example: Continuing with the previous example, suppose the covariance between the risky asset and the growth rate in the fund's liabilities is 0.002, and the value of liabilities is 20% higher than the value of assets. What will be the optimal weight of the equity allocation?

$$w = \frac{1}{10}\frac{0.09 - 0.01}{0.13^2} + 1.2\frac{0.002}{0.13^2} = 61.5\%$$

It can be seen that, compared to the previous example, the fund will hold about 14% more in the risky asset because it can hedge some of the liability risk.

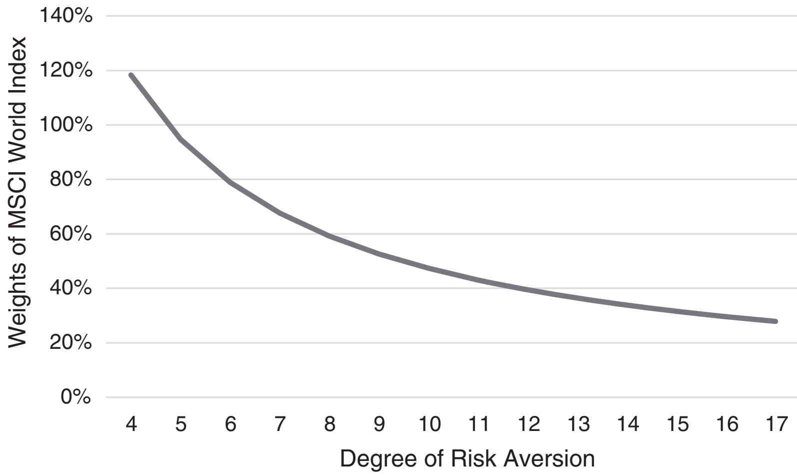


EXHIBIT 1.5 Optimal Weights of Risky Investment and Degree of Risk Aversion

By changing the degree of risk aversion in the first example, we can obtain a set of optimal portfolios, as shown in Exhibits 1.5 and 1.6.

It can be seen that at low degrees of risk aversion (e.g., 4), the investor will be investing more than 100% in the MSCI World Index, which means a leveraged position will be used. In addition, we can see the full set of expected returns and volatility that the optimal portfolios will assume, which is referred to as the efficient frontier.

The points appearing in Exhibit 1.6 correspond to various degrees of risk aversion. For instance, the optimal risk-return trade-off for an investor with a degree of

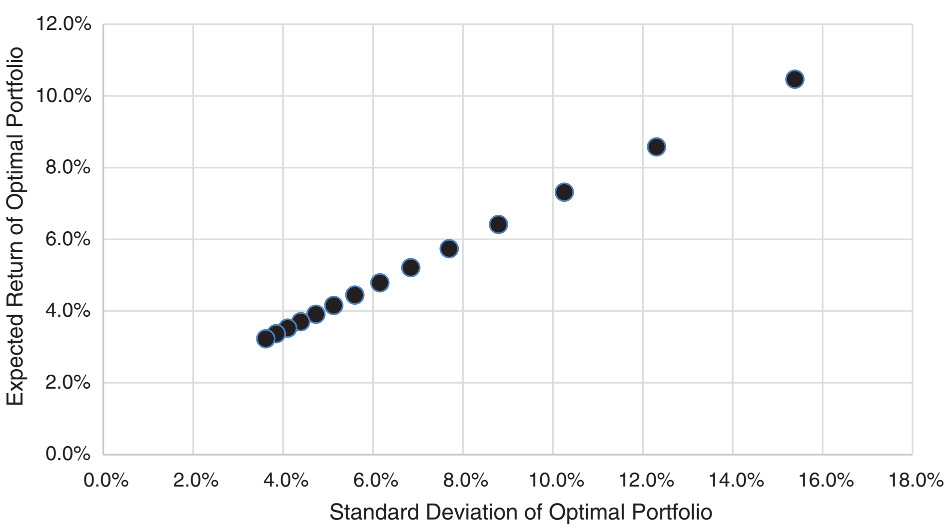


EXHIBIT 1.6 Expected Returns and Standard Deviations of Optimal Portfolios

risk aversion of 4 is represented by a portfolio that is expected to earn 10.5% with a volatility of about 15.4%.

1.8.4 Mean-Variance Optimization with Multiple Risky Assets

It turns out that a similar graph will be obtained even if the number of asset classes is greater than one. In that case, the graph will be referred to as the efficient frontier. The **efficient frontier** is the set of all feasible combinations of expected return and standard deviation that can serve as an optimal solution for one or more risk-averse investors. Put differently, no portfolio can be constructed with the same expected return as the portfolio on the frontier but with a lower standard deviation, or, conversely, no portfolio can be constructed with the same standard deviation as the portfolio on the frontier but with a higher expected return.

Example: In this example, the set of risky asset classes is expanded to three. The necessary information is provided in Exhibit 1.7. The figures are estimated using monthly data in terms of USD. The annual riskless rate is assumed to be 1%. Note that these estimates are typically adjusted to reflect current market conditions. This example is meant to illustrate an application of the model.

Using the optimal solution that was displayed in Equation 1.18, the optimal weights of a portfolio consisting of the three risky assets and one riskless asset can be calculated for different degrees of risk aversion. The results are displayed in Exhibit 1.8.

A number of interesting observations can be drawn from these results. First, notice that the optimal weights are not very realistic. For example, for every degree of risk aversion, the optimal portfolio requires us to take a short position in the MSCI World Index. Second, the optimal investment in the HFRI index exceeds 100% for some degrees of risk aversion considered here. Third, unless the degree of risk aversion is increased beyond 40, the optimal portfolio requires some degree of leverage (i.e., negative weight for the Treasury bills). Finally, the bottom two rows display annual mean and annual standard deviation of the optimal portfolios. These represent points on the efficient frontier.

EXHIBIT 1.7 Statistical Properties of Three Risky Asset Classes

			Annual Variance-Covariance		
1990–2015	Annual Mean	Annual Standard Deviation	MSCI World	Barclays Global Aggregate	HFRI Fund Weighted Composite
MSCI World	8%	15%	0.0234	0.0025	0.0073
Barclays Global Aggregate	6%	5%	0.0025	0.0029	0.0006
HFRI Fund Weighted Composite	10%	7%	0.0073	0.0006	0.0046

Source: Bloomberg and authors' calculations.

EXHIBIT 1.8 Optimal Weights and Statistics for Different Degrees of Risk Aversion

	Degree of Risk Aversion			
	10	**15**	**20**	**40**
MSCI World	−86%	−57%	−43%	−22%
Barclays Global Aggregate	186%	124%	93%	46%
HFRI Fund Weighted Composite	307%	205%	153%	77%
Treasury Bills	−307%	−171%	−103%	−2%
Optimal Portfolio Statistics				
Annual Mean	31%	21%	16%	9%
Annual Standard Deviation	17%	12%	9%	5%

Source: Authors' calculations. (Note that because of rounding errors, the weights do not add up to one.)

As we just saw, mean-variance optimization typically leads to unrealistic weights. A simple way to overcome this problem is to impose limits on the weights. For instance, in the example just provided, we can impose the constraint that the weights must be nonnegative. Unfortunately, when constraints are imposed on the weights, a closed-form solution of the type presented in Equation 1.18 can no longer be obtained, and we must use a numerical optimization package to solve the problem.[6]

If we repeat the example but impose the constraint that weights cannot be negative, the resulting optimal portfolios will reflect those displayed in Exhibit 1.9.

It can be seen that the weight of the MSCI World Index is constantly zero for all degrees of risk aversion. This means that portfolios that are on the efficient frontier in this case do not have any allocation to the MSCI World Index. Another important point to consider is that the optimal portfolios do not have the same attractive risk-return properties. By imposing a constraint, the resulting portfolios are not as optimal as they were when there were no constraints.

The mean-variance optimization approach discussed in this section can be presented in different forms. The Appendix at the end of this book provides two alternative methods that have appeared in the literature. The advantages of the approach presented in this section are twofold. First, as Equations 1.18 and 1.19 showed,

EXHIBIT 1.9 Optimal Weights without Short Sale Constraints

	Degree of Risk Aversion			
	10	**15**	**20**	**40**
MSCI World	0%	0%	0%	0%
Barclays Global Aggregate	138%	92%	69%	34%
HFRI Fund Weighted Composite	176%	117%	88%	44%
Treasury Bills	−214%	−109%	−57%	22%
Optimal Portfolio Statistics				
Annual Mean	23%	16%	12%	7%
Annual Standard Deviation	15%	10%	7%	4%

simple closed-form solutions can be obtained when there are no constraints. Second, the approach can be easily expanded to take into account preferences for higher moments of the probability distributions of asset returns. This will turn out to be important for our purpose, as alternative investments tend to have return distributions that deviate significantly from the normal distribution; therefore, their higher moments will be of interest to investors.

1.8.5 Hurdle Rate for Mean-Variance Optimization

An interesting implication of mean-variance portfolio optimization when there is a riskless asset is that the benefits of diversification can be shown to cause low-return assets to be desirable for inclusion in a portfolio. It is easy to show that the addition of a new asset to an already optimal portfolio will improve its risk-return properties (i.e., increases the expected utility) if the expected rate of return on this new asset exceeds a hurdle rate. The **hurdle rate** is an expected rate of return that a new asset must offer to be included in an already optimal portfolio. An asset being considered for addition to a portfolio should be included in the portfolio when the following expression is satisfied:

$$E[R_{\text{New}}] - R_f > (E[R_P] - R_f) \times \beta_{\text{New}} \tag{1.21}$$

Here, $E[R_{\text{New}}]$ is the expected rate of return on the new asset, $E[R_p]$ is the expected rate of return on the optimal portfolio, R_f is the riskless rate, and β_{New} is the beta of the new asset with respect to the optimal portfolio.[7]

Equation 1.21 states that the addition of the new asset to an optimal portfolio will improve the risk-return properties of the portfolio if the expected excess rate of return on the new asset exceeds the expected excess rate of return on the optimal portfolio times the beta of the new asset. If the new asset satisfies Equation 1.21, then its addition to the optimal portfolio will move the efficient frontier in the northwest direction. For example, if the beta of the new asset is zero, then the new asset will improve the optimal portfolio as long as its expected rate of return exceeds the riskless rate. If the new asset has a negative beta, then it could improve the optimal portfolio even if its expected rate of return is negative. In other words, assets that can serve as hedging instruments could have negative expected returns and still improve the performance of a portfolio.

APPLICATION 1.8.5

Suppose that an investor is using mean-variance optimization, the expected annual rate of return of an optimal portfolio is 16%, and the riskless rate is 1% per year. What is the hurdle rate for a new asset that has a beta of 0.5 with respect to the optimal portfolio?

Given the formula of Equation 1.21, the hurdle rate would be:

$$E[R_{\text{New}}] - 1\% > (16\% - 1\%) \times 0.5 \quad [R_{\text{New}}] > 8.5\%$$

What if the new asset has a beta of –0.3, which means that it can hedge some of the portfolio's risk?

$$E[R_{\text{New}}] - 1\% > (16\% - 1\%) \times (-0.3) \qquad E[R_{\text{New}}] > -3.5\%$$

In this case, even if the new asset is expected to lose some money (i.e., less than 3.5%), its addition to the optimal portfolio could still improve its risk-return properties.

1.8.6 Issues in Using Optimization

We have already seen that even in the case of three risky asset classes and the riskless rate, reasonable estimates of the weights could not be obtained unless short sale restrictions were imposed, and even in that case, no allocation to the MSCI World Index was recommended. This lack of allocation to an asset class was driven mostly by the portfolio's high volatility and relatively low return. In other words, using the past as an unbiased forecast of the future and using global equities as an asset class would have meant no allocation would be made to equities, and the portfolio would have focused on the remaining two assets. In practice, implementing an optimization method (mean-variance optimization, in particular) for portfolio allocation decisions raises major challenges.

1.8.7 Optimizers as Error Maximizers

Portfolio optimizers are powerful tools for finding the best allocation of assets to achieve superior diversification, given accurate estimates of the parameters of the return distributions. When the mean-variance method is used, there is a need for accurate estimates of expected returns and the variance-covariance matrix of asset returns. However, portfolio optimizers that use historical estimates of the return distributions have been derogatorily called "error maximizers" due to their tendency to generate solutions with extreme portfolio weights. For example, very large portfolio weights are often allocated to the assets with the highest mean returns and lowest volatility, and very small portfolio weights are allocated to the assets with the lowest mean returns and highest volatility. It is then argued that assets with the highest estimated means are likely to have the largest positive estimation errors, whereas assets with the lowest estimated means are likely to have the largest negative estimation errors. Hence, mean-variance optimization is likely to maximize errors. Therefore, if an analyst overstates mean returns and understates volatility for an asset, then the weights that the model recommends are likely to be much larger than an institutional investor would consider reasonable. Further, other assets are virtually omitted from the portfolio if the analyst supplies low estimates of mean returns and high estimates of volatility.

A typical attempt to use a mean-variance optimization model for portfolio allocation is this: (1) The portfolio manager supplies estimates of the mean return, volatility, and covariance for all assets; (2) the optimizer generates a highly unrealistic solution that places very large portfolio weights on what are considered the most attractive assets, with high mean return and low volatility, and zero or minuscule portfolio weights on what are considered the least attractive assets, with low mean return and high volatility; and (3) the portfolio manager then modifies the model by adding constraints or altering the estimated inputs—including mean, variance, and covariance—until the resulting portfolio solutions appear reasonable.

The problem with this process is that the portfolio weights become driven by the subjective judgments of the analyst rather than by the analyst's best forecasts of risk and return. The remaining sections discuss a variety of challenges that emanate from the tendency of mean-variance portfolio optimizers to select extreme portfolio weights.

It is important to point out that while higher-frequency data tends to improve the accuracy of the estimated variance-covariance matrix, it will do nothing to improve the accuracy of the estimated means; only a longer history has the potential to do so. To see this, assume that we have five years of annual data on the price of an asset. The annual rate of return on the asset is calculated to be $R_{t+1} = \ln(P_{t+1}/P_t)$. Now consider an estimate of the average return using the observed four annual returns:

$$\bar{R} = \frac{1}{4}\{\ln(P_2/P_1) + \ln(P_3/P_2) + \ln(P_4/P_3) + \ln(P_5/P_4)\} = \frac{1}{4}\{\ln(P_5/P_1)\} \tag{1.22}$$

Notice that all the intermediate prices cancel out, and only the first and the last prices matter. This result will not change even if one could use daily or even high-frequency data. The accuracy of the mean depends on the length of data and not on the frequency of the observations.

This observation regarding mean accuracy leads to the following dilemma. To obtain accurate estimates of the mean, it is necessary to have a very long history of prices. However, firms, industries, and economies go through drastic changes over long periods, and it would be highly unlikely that all observed prices would have come from the same distribution. In other words, of all the estimated parameters, the estimated mean is most likely to be the least accurate, yet it is the one with the most influence on the outputs of the mean-variance optimization.

The final difficulty in deriving estimates of return and risk for each asset class is that return and risk are nonstationary, meaning that the levels of risk and return vary substantially over time. Therefore, the true risk and return over one period may be substantially different from the risk and return of a different period. Thus, in addition to traditional estimation errors for a stationary process, estimates for security returns may include errors from shooting at a moving target.

1.8.8 Data Issues for Illiquid Assets

As noted, mean-variance optimizers can be error maximizers. Therefore, erroneous forecasts of the mean, variance, and covariance can result in extreme portfolio weights, with a resulting portfolio concentration in a few assets with estimated high means and estimated low volatilities. Most institutions view such concentrated positions as unacceptable speculation on the validity of the forecasted mean and volatility.

Although higher frequency of observed data can improve the accuracy of the estimated variance and covariance, for most alternative assets, high-frequency data is not available. More important, the assets whose prices cannot be observed with high frequency tend to be illiquid, and the reported quarterly returns are based on appraisals such as those used in real estate and private equity. These prices tend to be smoothed and therefore can substantially understate the variance and covariance of returns. Because volatility and covariance are key inputs in the optimization process, asset classes with low estimated correlation and volatility receive relatively large weights in the optimal portfolio. If smoothing has caused the reported volatility and correlation of an asset to substantially underestimate the true volatility, then a traditional mean-variance optimizer would overweight the asset. In this case, and to prevent extremely large allocations to assets with smoothed returns, the time series of returns may be unsmoothed, as discussed in Chapter 15, before being added to the optimization routine. But unsmoothing is imperfect, and other issues with accurately forecasting volatility and correlation remain.

1.8.9 Data Issues for Large-Scale Optimization

The problems with covariance estimation include the potentially large scale of the inputs required. This is typically not an issue when working at the asset class level, at which the investor may consider 10 asset classes for inclusion in the portfolio. However, optimizing an equity portfolio selected from a universe of 500 stocks has very large data requirements. A 500-asset optimization problem requires estimates of covariance between each pair of the 500 assets. Not only does this problem require $n(n-1)/2$, or 124,750 covariance estimates, but it is also difficult to be confident in these estimates, especially when there are too many to analyze individually. Also, notice that to estimate 124,750 covariance terms, we need more than 124,750 observations, or more than 10,000 years of monthly data or more than 340 years of daily data.

The problem of needing to calculate thousands of covariance estimates can be reduced with factor models. Rather than estimating the relationships between each pair of stocks in a 500-stock universe, it can be easier to estimate the relationship between each stock and a limited number of factors. While some investors simply choose to estimate the single-factor market model beta of each stock, others use multifactor models. To see how a factor model can reduce the data requirement, suppose the return on each asset can be expressed as a function of one common factor and some random noise:

$$R_i = a_i + b_i F + \varepsilon_i \tag{1.23}$$

where F is the common factor and a_i, b_i are the estimated parameters. It is assumed that for two different assets, the error terms ε_i and ε_j are uncorrelated with each other. Under this assumption, the covariance between two assets is given by:

$$\text{Cov}[R_i, R_j] = b_i \times b_j \times \text{Var}[F] \tag{1.24}$$

This means that to estimate the covariance matrix of 500 assets, we need 500 estimates of b_i and one estimate of $\text{Var}[F]$.

1.8.10 Mean-Variance Ignores Higher Moments

A problem that is especially acute with alternative investments is that the mean-variance optimization approach considers only the mean and variance of returns.

This means that the optimization model does not explicitly account for skewness and kurtosis. Investors' expected utility can be expressed in terms of mean and variance alone if returns are normally distributed. However, when making allocations to alternative investments and other investments with nonzero skewness and nonzero excess kurtosis, portfolio optimizers tend to suggest portfolios with desirable combinations of mean and variance but with highly undesirable skewness and kurtosis. In other words, although mean-variance optimizers can identify the efficient frontier and help create portfolios with the highest Sharpe ratios, they may be adding large and unfavorable levels of skewness and kurtosis to the portfolio.

For example, two assets with returns that have the same variance may have very different skews. In a competitive market, the expected return of the asset with the large negative skew might be substantially higher than that of the asset with the positive skew to compensate investors willing to bear the higher downside risk. A mean-variance optimizer typically places a much higher portfolio weight on the negatively skewed asset because it offers a higher mean return with the same level of variance as the other asset. The mean-variance optimizer ignores the unattractiveness of an asset's large negative skew and, in so doing, maximizes the error.

There are three common ways to address this complication. First, as we saw earlier, it is possible to expand our optimization method to account for skewness and kurtosis of asset returns. Second, we can continue with our mean-variance optimization but add the desired levels of the skewness and kurtosis as explicit constraints on the allowed solutions to the mean-variance optimizer, such as when the excess kurtosis of the portfolio returns is not allowed to exceed 3, or when the skewness must be greater than –0.5. A problem with incorporating higher moments in portfolio optimization is that these moments are extremely difficult to predict, as they are highly influenced by a few large negative or positive observations. In addition, in the second approach, a portfolio with a desired level of skewness or kurtosis may not be feasible at all. Finally, the analyst may choose to explicitly constrain the weight of those investments that have undesirable skew or kurtosis. For example, the allocation to a hedge fund strategy that is known to have large tail risk (e.g., negative skew) might be restricted to some maximum weight.

1.8.11 Other Issues in Mean-Variance Optimization

The results from a mean-variance optimization can be extremely sensitive to the assumptions, as small changes in the mean return or covariance matrix (i.e., the set of all variances and covariances) can lead to enormously different prescribed portfolio weights. The high sensitivity of portfolio optimizers to the input data has led to approaches that attempt to harness the power of optimization to identify diversification potential without generating extreme portfolio weights. In addition, in most cases, portfolio managers want to adjust the historical estimates to reflect their views about the estimated parameters going forward. For instance, a portfolio manager may want to incorporate her view that the health care sector is likely to do better than indicated by its historical track record. Perhaps the most popular modification to account for views and obtain reasonable estimates of weights is described by Black and Litterman.

The first problem addressed by the Black-Litterman approach is the tendency of the user's estimates of mean and variance to generate extreme portfolio weights in a mean-variance optimizer. Note that if a security truly and clearly offered a large expected return, low risk, and high diversification potential, then demand for the security would drive its price upward and its expected return downward until the demand for the security equaled the quantity available. In competitive markets, securities prices tend toward offering a perceived combination of risk and return in line with other assets.

The key to understanding the Black-Litterman approach is to understand that if a security offers an equilibrium expected return, then the demand for the asset will equal the supply. Further, the optimal allocation of the asset into every well-diversified portfolio will be equal to the weight of the asset in the market portfolio (i.e., the market weight). Thus, an equilibrium expected return for a security is the expected return that causes the optimal weight of that security in investor portfolios to equal its market weight.

This observation means that if the portfolio manager has no views about the future performance of a particular asset class, then its market weight should be used. For instance, a market-cap-weighted portfolio of global equities would be optimal. However, since market cap weights are not well defined for some asset classes, the Black-Litterman approach will need to be adjusted for application to alternative assets.

The primary innovation of the Black-Litterman approach is that it allows the investor to blend asset-specific views of each asset's expected return with views that would be consistent with market weights in a market equilibrium model.

Whereas some asset allocators employ advanced techniques such as the Black-Litterman approach to reduce the sensitivity of the weights to the expected risks and returns, a much larger number of asset allocators choose to add additional constraints to the optimization model to circumvent the difficulties and sensitivities of mean-variance optimization. Common additional constraints include the following:

- Limits on estimated correlation between the return on the optimal portfolio and the return on a predefined benchmark
- Limits on divergences of portfolio weights from benchmark weights
- Limits or ranges on the prescribed portfolio weights

The last constraint, limits on portfolio weights, is the most popular. These constraints can prescribe upper or lower weights, outside of which the asset allocator will not invest. The portfolio optimizer is forced to generate weights within those ranges. In practice, however, many investors use so many constraints that the constraints have more influence on the final asset allocation than does the mean-variance optimization process. While each of the added constraints may help the asset allocator avoid extreme weights, the approach may ultimately lead to having the constraints define the allocation rather than the goal of diversification.

1.9 CONCLUSION

This chapter has introduced the asset allocation process, with a focus on using the mean-variance approach to create optimal portfolios. The asset allocation process discussed in this chapter consists of five steps, of which four were discussed.

Step 1 focuses on understanding who the asset owners are and their mission in managing assets. Step 2 examines the asset owner's objectives and constraints. Here we discussed the expected utility and its mean-variance version as a flexible way of quantifying an asset owner's objectives. Two types of constraints, internal and external, were explained.

Step 3 deals with preparing the investment policy statement, which will provide a general framework for the actual asset allocation. One of the key features of this statement is to develop a list of asset classes to be considered. Step 4 is implementation, which was covered with a focus on mean-variance optimization and its potential problems.

NOTES

1. Hundreds of studies have attempted to determine if active managers outperform passive strategies. S&P Dow Jones Indices publishes SPIVA® U.S. Scorecard on a regular basis. It reports on the relative performance of U.S. mutual funds.
2. More detailed discussions of asset allocation processes can be found in Maginn et al. (2007) and Ang (2014).
3. For further details, see Maginn et al. (2007).
4. For a detailed discussion of strategic asset allocation, see Eychenne, Martinetti, and Roncalli (2011) and Eychenne and Roncalli (2011).
5. In a simple equilibrium model, the short-term real riskless rate is shown to equal the real growth in the economy minus a premium that depends on the volatility of the economy's real growth rate and the degree of risk aversion. See Cox, Ingersoll, and Ross (1985).
6. The problem is still a rather standard optimization program and can be solved using Solver from Excel or similar packages.
7. From linear regression and the CAPM we know that $\beta_{\text{New}} = \text{Cov}[R_P, R_{\text{New}}]/\text{Var}[R_p]$.

REFERENCES

Ang, A. 2014. *Asset Management: A Systematic Approach to Factor Investing*. Oxford: Oxford University Press.

Brinson, G. P., L. R. Hood, and G. L. Beebower. 1986. "Determinants of Portfolio Performance." *Financial Analysts Journal* 42 (4): 39–44.

Cox, J., J. Ingersoll, and S. A. Ross. 1985. "An Intertemporal General Equilibrium Model of Asset Prices." *Econometrica* 53 (2): 363–84.

Eychenne, K., S. Martinetti, and T. Roncalli. 2011. "Strategic Asset Allocation." Lyxor Asset Management, Paris.

Eychenne, K., and T. Roncalli. 2011. "Strategic Asset Allocation—An Update Following the Sovereign Debt Crisis." Lyxor Asset Management, Paris.

Ibbotson, R. G., and P. D. Kaplan. 2000. "Does Asset Allocation Policy Explain 40, 90, or 100 Percent of Performance?" *Financial Analysts Journal* 56 (1): 26–33.

Maginn, J. L., D. L. Tuttle, J. E. Pinto, and D. W. McLeavey. 2007. *Managing Investment Portfolios*. 3rd ed. Hoboken, NJ: John Wiley & Sons.

CHAPTER 2

Tactical Asset Allocation, Mean-Variance Extensions, Risk Budgeting, Risk Parity, and Factor Investing

Chapter 1 discussed the asset allocation process, strategic asset allocation, and the basic mean-variance approach. The role of the investment policy statement was examined as a way of summarizing the objectives and constraints of asset owners, and strategic asset allocation was presented as a long-term optimal allocation. Next, the basic properties of the mean-variance approach were examined as a quantitative method for creating optimal portfolios that are consistent with asset owners' objectives. This chapter begins with a discussion of the methodology behind tactical asset allocation (TAA), and then studies some practical extensions to the mean-variance model that address some of the problems raised at the end of Chapter 1. In particular, it examines how illiquidity, risk factor exposures, and estimation risks can be taken into account when creating optimal portfolios. This chapter then discusses alternatives to the basic mean-variance optimization. It discusses risk budgeting as a way of understanding and controlling the risk exposures of a portfolio; it also examines the risk parity approach, which is closely related to risk budgeting. Finally, it presents a relatively new approach called factor investing, which recommends an optimal risk allocation rather than an optimal asset allocation as the proper way of creating optimal portfolios.

2.1 TACTICAL ASSET ALLOCATION

Tactical asset allocation has a long history and has been used by large and small asset owners.[1] However, there is no well-established definition of TAA, and different authors and investment firms use the term to mean different things. One thing that all uses of TAA seem to have in common is that it represents a form of active management of a portfolio. Tactical asset allocation (TAA) is defined as an active strategy that shifts capital to those asset classes that are expected to offer the most attractive risk-return combination over a short- to medium-term time horizon. In this sense, TAA is a dynamic asset allocation strategy that actively adjusts a portfolio's strategic asset allocation (SAA) based on short- to medium-term changes in the economic and financial environment. TAA will add value if it can systematically take advantage of temporary market inefficiencies and departures of asset prices from their fundamental values. As discussed in Chapter 1, over time, SAA allocation is the most important driver of a portfolio's risk-return characteristics. TAA can add value if (1) there are short- to medium-term inefficiencies in some markets, and (2) a systematic approach can be designed to exploit these inefficiencies while overcoming the risks and costs that are associated with active portfolio management.[2]

2.1.1 TAA and the Fundamental Law of Active Management

Chapter 20 of the CAIA Level I book discussed the Fundamental Law of Active Management (FLOAM). This model expresses the risk-adjusted value added by an active portfolio manager as a function of the manager's skill to forecast asset returns and the number of markets to which the manager's skill can be applied (breadth). In particular, we saw that

$$IR = IC \times \sqrt{BR} \tag{2.1}$$

where *IR* is the information ratio and is equal to the ratio of the manager's alpha (i.e., expected outperformance) divided by the volatility of the alpha. *IC* is the information coefficient of the manager, which is a measure of the manager's skill, and represents the correlation between the manager's forecast of asset returns and the actual returns to those assets. *BR* is the strategy's breadth, which is defined as the number of independent forecasts that the manager can skillfully make during a given period of time (e.g., one year). Not surprisingly, the value added by active management increases with the ability of the manager to forecast returns and the number of independent markets to which the forecasting skill can be applied.

The FLOAM can be applied to security selection as well as asset class allocation. Clearly, when FLOAM is applied to the process of selecting securities from a universe of 5,000 or more, there should be greater potential for adding value, as the breadth could be large. This insight has been used as an argument against TAA. In other words, to add value through active asset allocation, a portfolio manager will need a much higher level of skill if that skill is to be applied to only a handful of independent asset classes.

APPLICATION 2.1.1

Suppose active manager A has the skill to select stocks from a universe of 2,000 securities and generate an information ratio (IR) of 1.2. This means that the expected alpha of this manager's portfolio is 20% higher than the volatility of the alpha. Active manager B can generate the same information ratio (1.2) using 15 asset classes. What are the managers' information coefficients?

Using Equation 2.1:
Active manager A has an information coefficient of 0.027:

$$1.2 = IC \times \sqrt{2{,}000} \qquad IC = 0.027$$

Active manager B has an information coefficient of 0.310:

$$1.2 = IC \times \sqrt{15} \qquad IC = 0.310$$

In this example, active manager B has to be about 11.5 times more skillful using active management among asset classes than active manager A, who is using security selection to achieve the same IR. In Equation 2.1, any of the variables can be solved given the value of the other two.

An extension of FLOAM provides additional insights into the potential value-added properties of TAA. FLOAM assumes that the manager is unconstrained in the sense that she can apply her skills to all available securities. In reality, portfolio managers face a number of constraints, both internal and external. For instance, the manager is constrained by the limits imposed by SAA. In addition, there could be regulatory constraints on allocations to certain asset classes. Finally, there are implementation costs associated with active management. This is particularly important when considering alternative investments. The costs associated with changing allocations to private equity or some real assets could be prohibitive. Even altering allocations to more liquid segments of alternative investments, such as commodity trading advisers (CTAs) and some hedge fund strategies, could be costly. The next section models potential costs.

2.1.2 FLOAM and the Cost of Active Management of Alternatives

Unlike traditional asset classes, most alternative assets are not actively traded, whether they are held directly or held through investment pools. Even investment pools with liquid underlying assets (e.g., CTAs and long/short equity hedge funds) are typically not liquid. Therefore, there are costs associated with actively managing portfolios that consist of alternative assets. This section introduces implementation costs into FLOAM.

The FLOAM as expressed in Equation 2.1 assumes that the manager faces no constraints in the asset allocation decision. If one were to take into account that some allocations may have to be substantially different from the ideal allocation recommended by the manager's forecasting skill, an extended FLOAM can be rewritten as:

$$IR = IC \times \sqrt{BR} \times TC \tag{2.2}$$

where *TC* is the transfer coefficient. TC measures the ability of the manager to implement her recommendations.[3] The upper limit for TC is one, and the lower limit is zero. When TC is equal to one, it means the manager is able to implement all her recommendations. Clearly, TC will be less than one when a portfolio of alternative asset classes is considered. Both hard constraints, such as no short selling of funds, and soft constraints, such as high transaction costs associated with the rebalancing of alternative assets, will reduce the value of TC well below one.

APPLICATION 2.1.2

Consider two similarly skilled active managers, each with an IC of 0.50. Active manager C can apply his skills to only 16 asset classes, whereas active manager D can apply her skills to 100 securities. What level of transfer coefficient does each manager need to have in order to generate an information ratio of 1.2?

Using Equation 2.2:

Active Manager C:

$$1.2 = 0.5 \times \sqrt{16} \times TC \qquad TC = 0.60$$

Active Manager D:

$$1.2 = 0.5 \times \sqrt{100} \times TC \qquad TC = 0.24$$

By having greater breadth, active manager D can achieve an IR of 1.2 with a TC of only 0.24.

2.1.3 Costs of Actively Managing Portfolios with Alternatives

Either as a result of monitoring or stemming from the desire to actively manage a portfolio of traditional and alternative assets, an investor may decide to redeem a fund or change allocations among funds. Even if the manager lacks the skill to forecast returns, such reallocations could provide benefits if the monitoring process has unveiled certain concerns about the manager. For example, the manager may have deviated from his original strategy or the fund may have experienced significant outflows, reducing his ability to spend the resources needed to manage an institutional-quality fund. While there may be benefits to redemption or rebalancing, there are also associated costs. The most important costs in actively managing a portfolio of funds are forgone loss carryforward opportunities and costs associated with liquidation and reinvestment: dormant cash, opportunity losses, and slippage from transaction costs and market impact.

2.1.3.1 The Cost of Forgone Loss Carryforward The forgone loss carryforward is potentially borne by every investor in a fund with an asymmetric fee structure.

Forgone loss carryforward arises when an existing investor loses the fee benefits of owning a fund below its high-water mark. The cost to the investor results from a managerial decision to liquidate a fund. Because a manager collects performance fees only when net asset value (NAV) is above the most recent high-water mark at the end of the relevant accounting period, a manager who is **underwater** (i.e., whose net asset value is below the most recent high-water mark) does not accrue performance fees until a new high-water mark is achieved.

The cost of loss carryforward should be taken into account when the decision is being made to replace a poorly performing manager with another manager. Going forward, the return realized from the poorly performing manager will be gross of performance fees, while the return earned on the investment with the new manager will be net of performance fees. This means the new manager will need to outperform the old manager by the amount of the performance fee just to break even. If the drawdown of the current manager is large, investors collect gross of fee returns for several periods. For example, if a manager experiences a drawdown of 25%, then the next 33.3% return ($0.75 \times 1.333 = 1.0$) generated by the manager is passed on to investors gross of performance fees. Assuming a performance fee of 20%, the new manager has to earn 41.67% [$0.333/(1 - 0.2) = 0.4167$] for the investor just to break even.

While the loss carryforward represents a potential cost for replacing a manager that has recently experienced some losses, there are four reasons that an investor may still wish to replace a manager with a carryforward loss. First, the investor may be concerned that the manager does not have an adequate incentive to generate performance until the high-water mark is reached. That is, because of the lack of incentive on the part of the manager, the recent poor performance may continue for some time. Second, in the absence of a performance fee, the management fee alone may not be enough for the fund to retain its best traders and maintain its risk management and compliance infrastructure. Therefore, the fund may no longer represent an institutional-quality fund. Third, other investors may withdraw their funds, making the investor's relative position in the fund too large. Most investors want to avoid this situation, because if they decide to redeem their shares in the future, the fund's NAV and operations could be adversely affected when a relatively large part of the assets under management (AUM) is redeemed. Finally, the investor may wish to reallocate away from the poorly performing manager because he believes the fund's strategy is no longer attractive. Whether based on the investor's decision to replace a manager or the manager's decision to liquidate the fund, the investor suffers a historical loss from paying incentive fees on profits that were lost and a prospective loss from not being able to earn future profits without paying incentive fees while the fund returns to its net asset value.

2.1.3.2 Four Other Costs of Replacing Managers One important cost associated with replacing a poorly performing manager was discussed in the previous section. There are four other costs associated with replacing managers, none of which are affected by the past performance of the managers. These include forgone interest on dormant cash, forgone excess returns on uncommitted cash, administrative costs of closing out one position and opening another, and the market impact of liquidating one position and starting a new position. The common factor driving the first two costs is that there are several leads and lags when making a decision to take money from one manager and placing it with another. The last two costs depend on the strategy being considered and the experience and resources of the investor.

To understand the first two costs (i.e., forgone interest on dormant cash and forgone excess returns on uncommitted cash), consider the lags that exist in the process of replacing one manager with another. The first lag represents the time it takes to review a manager's results and make a decision about redemption. This might take anywhere from a few days to, in some complex cases, several months. The second lag represents the time between the notification deadline for a withdrawal and the moment when a net asset value is struck. This can take several weeks. The third lag is the time between the striking of net asset value and the receipt of the first round of cash. The fourth lag represents the time that passes between the receipt of the first round of cash and the final round of cash. The last lag is related to the time between when the entire position is liquidated and the cash is returned to the investor and when the cash can be allocated to a new manager.

The first cost arises from the third and fourth lags and is associated with the liquidation and the forgone interest on dormant cash. This cost is borne by the investor, and depends entirely on a fund's practices with regard to interest payments on cash balances. Industry practices vary a great deal, but it is not uncommon to find that funds do not pay interest on the value of cash balances. In these cases, the cost of forgone interest depends on how quickly cash is returned to the investor.

The second cost is associated with the last lag and represents the opportunity losses associated with liquidation and reinvestment. These losses stem from the intervals during which investments are not committed to enterprises that promise returns in excess of market interest rates.

The third cost is related to transaction and administrative fees. Closing out old positions and opening new positions will entail administrative fees and due diligence costs. These costs will vary by investor. Experienced investors may have long lists of managers to choose from, and therefore due diligence costs may be relatively small. Also, since due diligence is a relatively fixed cost, the impact will depend on the size of the position. That is, this could be a significant cost for investors who have a relatively small allocation to this asset class.

The final cost will depend on the liquidity of the positions that have to be closed and opened. If the positions are not liquid or the strategy being considered does not have a large capacity, then the market impact of liquidating the old positions and creating new positions could be costly, and will be borne primarily by the investor.

2.1.4 Three Observations on TAA and Portfolio Reallocation Costs

Given the preceding discussion, it appears that it will be very hard to make a convincing case for tactical asset allocation using alternative asset classes. However, a few observations may provide a basis for tactical asset allocation.

First, by focusing on a few asset classes, the manager might be able to develop separate forecasting models for each and therefore generate forecasts with independent errors. While FLOAM presents *IC* and *BR* as somewhat independent parameters, they tend to be dependent in practice. For example, it is highly unlikely for a manager to have the skill to forecast returns on a large number of independent securities. Notice that the key word here is *independent*. This means that the manager applies one or more models to a set of securities that are not highly correlated, and therefore the forecast errors are independent from each other. This is a very strong requirement that is unlikely to be fully satisfied. In other words, there is a negative relationship between *IC* and *BR*. The more markets to which the manager tries to apply her skills, the less accurate the forecasts are likely to become.

Second, the information coefficient tends to be much higher when applied to asset classes than when applied to individual securities. The random returns on individual security prices contain a significant amount of noise, which makes forecasting models less accurate. On the other hand, available empirical evidence suggests that expected returns on various asset classes or portfolios of securities behave in a more predictable way through various market cycles.[4]

Finally, while TAA may be difficult and costly to apply to alternative assets, TAA can be applied to the traditional portion of the portfolio, where derivative products can be used to significantly alter a portfolio's characteristics without having to redeem allocations to certain illiquid funds. For example, suppose the equity beta of a diversified portfolio of traditional and alternative asset classes is given by β_{Port}. This can be estimated by regressing the historical returns of the portfolio against the returns on an equity benchmark. The portfolio manager can reduce the equity beta of this portfolio by selling equity futures contracts. In particular, since the beta of a portfolio is equal to the weighted average of the betas of its assets, we have:

$$\beta_{\text{New}} = \beta_{\text{Port}} + \frac{F}{P}\beta_{\text{Futures}} \tag{2.3}$$

Here, β_{New} is the new beta of the portfolio, which could serve as the target by the manager, (F/P) is the ratio of the notional amount of the positions in the futures contracts to the size of the portfolio, and β_{Futures} is the beta of the futures contract with respect to the equity benchmark used to calculate the beta of the portfolio, β_{Port}. The beta of the futures contract is typically equal to one. An investor can engineer a new beta for the portfolio by adjusting the level of futures contracts, *F*.

APPLICATION 2.1.4

Suppose the beta of a diversified portfolio against the S&P 500 Index is 0.9. The portfolio manager wants to increase the beta to 1.2 because of improving economic conditions. If the market value of the portfolio is \$500 million, what notional position does the portfolio manager need to take in the S&P 500 futures market with a beta of 1 in order to achieve a beta of 1.2? What position would have lowered the beta to 0.4? Inserting the known values into Equation 2.3:

$$1.2 = 0.9 + \frac{F}{500} \times 1 \qquad F = 150$$

That is, the manager has to take long positions in \$150 million worth of S&P 500 futures contracts. If the manager had decided to reduce the equity exposure to 0.4, the futures position would have been:

$$0.4 = 0.9 + \frac{F}{500} \times 1 \qquad F = -250$$

In the second case, the manager has to establish a \$250 million short position in S&P 500 futures contracts to lower the beta to 0.4 from 0.9.

Application 2.1.4 highlights an actual advantage that TAA may have when applied to broad asset classes: Portfolio managers may use liquid futures and swap markets to effectively implement the ultimate objective of TAA, which is to alter the portfolio's exposures to various sources of risk. This issue of factor exposures and factor investing will be further discussed later in the chapter.

2.1.5 Keys to a Successful TAA Process

As just illustrated, TAA has to overcome significant costs and barriers if it is to be applied to portfolios of alternative asset classes. However, TAA can also be applied to such a portfolio not through the actual sale and purchase of illiquid alternatives but through the use of futures contracts to alter the portfolio exposure to sources of risks. This section discusses the TAA process.

2.1.5.1 The TAA Process and Return Prediction The key component of a successful TAA process is the development of sound models that can consistently forecast returns across asset classes. This point may appear to contradict the basic tenet of the efficient market hypothesis that asset returns are essentially unpredictable. There is strong evidence that returns to asset classes are indeed predictable (Pesaran 2010). What is less clear is whether this predictability is the result of foreseeable changes in risk premiums or market inefficiency. It is beyond the scope of this chapter to discuss the academic and industry findings regarding the sources of predictability in asset returns. It suffices to say that studies have not been able to conclusively rule out that some asset return predictabilities using low-frequency data (e.g., annual frequency) represent temporary inefficiencies in markets (Asness, Moskowitz, and Pedersen 2010; Hull and Qiao 2015). The effectiveness of a TAA strategy is largely dependent on constructing a good model. As mentioned, the first step in developing a TAA strategy is to forecast excess returns by constructing fundamental and technical models that can predict asset class returns, using a set of explanatory variables for fundamental models and signals for technical models.[5]

Models may have varying predictive strengths during different economic regimes. In addition, according to the FLOAM, the potential value added of TAA is increased if the models' errors are not correlated with each other. Therefore, multiple forecasting models should be used to maximize the value added by the TAA strategy. A good forecasting model must include economically meaningful signals and have a research process that correctly identifies those signals. In addition, the model must have performed well in the past using out-of-sample data.

2.1.5.2 Three Notable Characteristics of Sound TAA Model Development The following are three important characteristics of sound model development:

1. Use of economically meaningful signals. **Economically meaningful signals** are those signals with rational, intuitive explanations for their expected predictive power. For example, the term spread, as an indicator of the business cycle, is intuitive and rational. Theoretically, a flat or downward-sloping yield curve is associated with a lower inflation rate and slower economic activity. A model that uses inputs with a strong basis in economic theory is likely to provide signals that identify fundamental shifts in the economy.
2. Absence of data mining. The manager should be able to confirm that the predictive results are not due to data mining. This could occur if the manager tries a variety of models and explanatory variables to see which one performs best and chooses the most accurate one to implement the TAA strategy. Ex post, it is possible to find explanatory variables that can predict the most random processes. The question is whether there are economic reasons to think that ex ante the model would have worked. More important, out-of-sample validations must be done to guard against data mining. Out-of-sample tests of the strategy, such as in other time periods or countries, can help confirm that the strategy's success is not simply the result of fitting the model to explain one historical period.
3. Avoidance of overfitting. Models that have a large number of explanatory variables can produce impressive r-squareds, especially when there is limited data. While models with smaller numbers of explanatory variables are likely to produce less impressive r-squareds using the same sample data, they are more likely to reproduce their in-sample predictive power out of sample. In addition, models that use fewer variables are more likely to be stable through time; that is, the estimated relationship does not change radically because of small changes in the data.

2.1.5.3 Fundamental Analysis Underlying TAA Models Linear regression is the most common approach for testing and developing fundamental models. Linear regressions are the simplest way to create conditional expectations of asset returns. Unlike unconditional expected returns, which use the average historical returns on the asset to form expectations, **conditional expectation models** obtain estimates of expected returns that are functions of the current values of a set of predictive variables.

Suppose we have collected a long time series of monthly returns on an emerging equity market, and we run the unusual regression in which the dependent variable is the equity return and the explanatory variable is the constant one. This regression has a single coefficient, the intercept, and it will be the unconditional expected return on the asset, and will be exactly equal to the historical average monthly return. The intercept may change slightly as new data is observed and added to the sample, but it will remain mostly unchanged, especially if a very long series is used.

EXHIBIT 2.1 S&P 500 Returns versus Lagged Dividend Yield

1919–2015	Value
Intercept	3.30%
Coefficient of Dividend Yield	1.04
R-squared	4%
Historical Means	
Annual Return on S&P 500	7.70%
Annual Dividend Yield	4.20%

Suppose the same asset returns are now regressed against the constant one, and lagged values of the term spread in the local bond market. The expected return on this emerging market conditioned on the term spread is now equal to the estimated intercept plus the estimated slope coefficient times the current value of the term spread. We can think of TAA as SAA when conditional expected returns are used as inputs in the asset allocation process. For example, in a mean-variance model, conditional expected returns are used to generate tactical weights. These weights may change drastically if there is a fundamental change in economic conditions leading to substantial changes in the conditioning variables.

For example, annual returns on the S&P 500 Index are regressed against lagged dividend yields on the S&P 500 using data from January 1919 to December 2015. The regressions results are presented in Exhibit 2.1.

As expected, this simple model lacks significant predictive power, as the slope coefficient, 1.04, is not statistically significant (the t-stat is 0.97). Still, it can be used to demonstrate how to employ such models. The unconditional mean return on the S&P 500 is 7.7% per year. To calculate the conditional mean, we need the current value of the dividend yield, which was about 2.11% in December 2015. Therefore, the conditional mean return on the S&P 500 in December 2015 is:

$$E[R_{\text{S\&P500}} | DivYld_{\text{Dec2015}}] = 3.3\% + 1.04 \times 2.11\% = 5.5\%$$

The conditional mean return on the S&P 500 for 2016 is estimated to be 5.5%. Should the dividend yield be different in December 2016, the conditional expectation of S&P 500 returns for 2017 will change accordingly.

2.1.5.4 Technical Analysis Underlying TAA Models While regression models based on fundamental economic relationships can be useful in implementing TAA strategies, models based on technical analysis can be used to supplement these models; in some cases, research has shown that they may even perform better than fundamental models. Faber (2013) has produced one of the most cited studies in this area. The approach employed in the paper is very simple but has produced results that seem to indicate that technical analysis could serve as the basis of TAA. The model uses six asset classes: U.S. large cap (S&P 500), non-U.S. developed markets (MSCI EAFE), U.S. 10-year government bonds, commodities (S&P GSCI), real estate

EXHIBIT 2.2 Global Tactical Asset Allocation

1973–2012	Buy and Hold	TAA
Annual Realized Returns	9.92%	10.48%
Annual Realized Volatility	10.28%	6.99%
Sharpe Ratio	0.44	0.73
Maximum Drawdown	–46.12%	–9.54%

Source: Based on results from Faber (2013).

investment trusts (NAREIT Index), and 90-day U.S. Treasury bills. There is an expanded version of the model that uses more asset classes.

The strategy examines a very simple quantitative TAA model, which is based on a trend-following model. The objective is to build a simple model that can be used to highlight the use of technical signals in performing TAA for traditional asset classes. The model is mostly a risk-reduction technique that signals when a portfolio manager should reduce exposure to a risky asset class in favor of a less risky investment. The model uses the following buy and sell rules:

- Increase exposure to 20% of the portfolio if the current price is above the 10-month simple moving average.
- Reduce exposure to 0% of the portfolio if the current price is below the 10-month simple moving average.

The model is updated monthly, and the cash not allocated to risky strategies is invested in Treasury bills. Exhibit 2.2 presents the results reported using data covering 1973–2012.

The buy-and-hold strategy is an equally weighted portfolio. The model has certainly performed well in the past. It is important to note that transaction costs and fees have not been taken into account. Therefore, it is reasonable to assume that the mean return would have been closer to 9.92% if costs had been taken into account. However, the risk reduction would not be impacted by these costs.

2.2 EXTENSIONS TO THE MEAN-VARIANCE APPROACH

This section examines some extensions to the mean-variance approach that have been developed to address some of the issues raised in Chapter 1. First, it discusses how illiquidity can be incorporated into the mean-variance model. Using a measure of illiquidity, we can adjust the model to penalize those assets that are illiquid. That is, everything else being equal, portfolio managers would prefer to invest in liquid assets; therefore, illiquid assets will be considered only if they provide some additional benefit. Second, we discuss how limits on factor exposures can be incorporated into a mean-variance approach. Finally, we address how estimation risk can be taken into account. **Estimation risk** refers to the risk that the estimated parameters that are used as inputs in the mean-variance approach could be different from the true values of those parameters. For example, in Chapter 1 of this book, expected returns on asset classes were shown to be quite inaccurate if a long series of data was not available. A method for incorporating the risk of misestimating the mean return of an asset class will also be discussed.

2.2.1 Adjustment of the Mean-Variance Approach for Illiquidity

Investors, especially after the 2008–9 global financial crisis, have become more concerned with illiquidity risk and its costs. Before discussing liquidity risk and its incorporation into the mean-variance framework, we need to define liquidity more precisely. There are two types of illiquidity risks: market liquidity risk and funding liquidity risk (Hibbert et al. 2009). **Market liquidity risk** arises when an event forces an investor to sell an asset that is not actively traded and there are a limited number of active market participants. Under such a circumstance, the price of the asset may have to be reduced significantly in order to bring a bid from market participants. The discount offered by the seller will have to be significantly higher during periods of financial stress. It is important to note that some apparently liquid assets could become illiquid during periods of economic stress. Certain mortgage-backed securities that appeared to be quite liquid became highly illiquid during the 2008–9 financial crisis. Two factors especially contributed to increased illiquidity of certain assets during the period of financial stress: lack of transparency in some products and the presence of asymmetric information between sellers and buyers. For example, during the 2008–9 financial crisis, investors (buyers) were reluctant to purchase complex mortgage-backed securities because (1) buyers could not value the pools of mortgages that acted as collateral for those securities, and (2) buyers assumed that sellers must know that the pools contained many nonperforming mortgage loans, which was why they were willing to sell them at such deep discounts. Many alternative investments—such as private equity, some real assets, and hedge funds with long lockups—could expose an investor to market liquidity risk.

Funding liquidity risk arises when the investor is unable to obtain financing, cannot roll over currently available debt, or lacks liquidity to meet capital commitments. Funding liquidity risk can arise in the context of private equity as well, but this time it is related to the inability of the investor to meet the general partner's capital calls. The two risks are related in the sense that by having a substantial amount of cash, one would be able to reduce both types of liquidity risk. An unlevered portfolio of publicly traded securities has relatively low market liquidity risk because the securities can be sold on the exchange at relatively low cost. Also, funding liquidity risk is absent because the portfolio is not levered, and therefore funding cannot be withdrawn by lenders.

In this section, the focus is on market liquidity risk and its incorporation into the mean-variance framework. The following presentation is kept simple to highlight the impact of liquidity risk on optimal allocations. In particular, it presents a model that modifies the mean-variance optimization framework by incorporating a liquidity penalty function, the purpose of which is to allow for an explicit, easily communicated, natural specification of liquidity preferences that works in conjunction with the standard mean-variance approach that was discussed in Chapter 1.[6] The **liquidity penalty function** reflects the cost of illiquidity and the preference for liquidity. By incorporating this penalty function into the traditional mean-variance optimization model, we can construct a framework for asset allocation involving illiquid assets.

In Chapter 1, the general framework of the mean-variance optimization was presented as:

$$\max \bar{R}_P - \frac{\lambda}{2}\sigma_P^2 \qquad \text{subject to } w_i \geq 0 \quad i = 1, \ldots, N \tag{2.4}$$

Here, $\bar{R}_P = \sum_{i=1}^{N} w_i(\bar{R}_i - R_0) + R_0$ is the expected rate of return of the portfolio, w_i is the weight of asset i in the portfolio, λ is the investor's measure of risk aversion, and σ_P^2 is the variance of the rate of return on the portfolio.

To account for illiquidity, we begin by assigning each asset class i an illiquidity level, denoted by L_i, which takes values between 0 and 1. Perfectly liquid assets are assigned 0 and highly illiquid assets are assigned 1. The liquidity level of the portfolio is measured by $L_P = \sum_{i=1}^{N} w_i L_i$. The objective function of the mean-variance approach is now adjusted to reflect a penalty for the extent that each feasible portfolio displays illiquidity when the investor has a preference for liquidity:

$$\max \quad \bar{R}_p - \frac{\lambda}{2}\sigma_P^2 - \phi L_P \qquad \text{subject to } w_i \geq 0 \quad i = 1, \ldots, N \tag{2.5}$$

Here, ϕ is a positive number and it represents the investor's preference for liquidity (i.e., aversion to illiquidity). If all assets are highly liquid, then $L_P = 0$ and the liquidity preference will have no impact. However, when some assets are illiquid, $L_P > 0$ and the investor's aversion to illiquidity will reduce the attractiveness of portfolios with illiquidity. The impact of the liquidity penalty will be to reduce the value of the objective function by subtracting the illiquidity penalty from the expected returns on illiquid assets. In other words, solving this problem would be similar to solving the original optimization problem of Equation 2.4, in which the expected return on each asset is reduced by ϕL_i. For the most illiquid assets, $L_i = 1$, which reduces the expected returns on such assets by ϕ. This insight can help the portfolio manager select a reasonable value for ϕ.

For example, a portfolio manager has assigned a liquidity level of 0.5 to the private equity asset class. The expected annual mean return on this asset class is estimated to be 18%. The asset owner is an endowment and therefore does not have strong preference for liquidity; this has led the portfolio manager to set $\phi = 0.10$. The adjustment to the mean return of the private equity asset class is:

$$\bar{R} - \phi L_i = 0.18 - 0.1 \times 0.5 = 13\%$$

The manager of a family office portfolio is considering the same asset class and assigns a liquidity level of 0.5 as well. However, liquidity is more important to this family office, and therefore the manager has set $\phi = 0.20$. The adjustment to the estimated expected return of the private equity asset class for the family office inside the model should be:

$$\bar{R}_i - \phi L_i = 0.18 - 0.2 \times 0.5 = 8\%$$

Obviously, everything else being equal, the family office will make a smaller allocation to the private equity class.

2.2.2 Adjustment of the Mean-Variance Approach for Factor Exposure

In the process of asset allocation, whether strategic or tactical, the investor may wish to limit the exposure of the portfolio to certain sources of risk (and potentially some returns) other than or in addition to the risk of the overall market. For example, the investor may wish to cap the exposure of the portfolio to changes in the price of oil. As long as an observable factor representing the source of risk exists, the constraints on factor exposures can be incorporated into the mean-variance approach with little difficulty.

The first step is to estimate the factor exposures of the assets that are being considered for inclusion in the portfolio. To do this, a linear regression of the following form is run:

$$R_{it} = a_i + b_i F_t + \varepsilon_{it} \tag{2.6}$$

Here, R_{it} is the rate of return on asset i; a_i and b_i are the intercept and the slope of the regression, respectively; F_t is the risk factor that is of interest; and ε_{it} is the residual. For example, F_t might be the percentage change in the price of oil. In this regression, b_i measures the factor exposure of asset i. Since the factor exposure of the portfolio is a weighted average of the individual asset's exposures, the factor exposure of the portfolio can be expressed as $b_P = \sum_{i=1}^{N} w_i b_i$. The following constraint can now be added to the mean-variance optimization of Equation 2.4:

$$b_P \leq \bar{b} \tag{2.7}$$

In this case, the constraint is that the total factor exposure of a potential portfolio must be less than or equal to the target, $\bar{b}$, in order for the portfolio to be a feasible solution. The impact of imposing this constraint is to reduce allocations to those asset classes that have large exposures to the source of risk being considered. It is important to note that since short sales are not allowed, it may not be feasible to create a portfolio with the desired level of factor exposure. For example, suppose an investor decides to have a negative exposure of the portfolio to the equity risk of the overall equity market. It may not be possible to create a negative equity beta portfolio when short positions are not permitted.

2.2.3 Adjustment of the Mean-Variance Approach for Estimation Risk

The inputs to the portfolio allocation process are unknown and have to be estimated. Further, accurate estimates of some inputs such as expected returns are more difficult to obtain. As seen in Chapter 1, to estimate the mean return on an asset class, a long price history is generally beneficial. This is particularly troublesome when considering alternative asset classes. Most of the funds that provide access to alternative assets have a relatively short history. As a result, the level of confidence in the estimated expected means and alphas of the funds is low. Therefore, it would be preferable if the portfolio optimization problem would be able to account for the uncertainty associated with the estimated input values. In particular, it seems reasonable to adjust the optimization process to reduce allocations to those asset classes that are subject to significant estimation risk (e.g., because a short return history is available).

Suppose an investor is considering an allocation to a relatively new hedge fund strategy for which only five years of monthly data are available (60 observations). Suppose the monthly mean and the standard deviation of an index representing the strategy are estimated to be 0.83% and 2.89%, respectively (equivalent to 10% annually for both). The estimated mean is a random variable, as one is likely to obtain different estimates when using different samples of returns, and we know from the CAIA Level I book that the standard error of the estimate mean will be $2.89\%/\sqrt{60}$. The Level I book discussed how confidence intervals can be calculated once the standard error of an estimate is known. In the present case, the 95% confidence interval for the estimated mean is approximately:

$$0.83\% - 1.96 \times \frac{2.89}{\sqrt{60}} \leq \bar{R} \leq 0.83\% + 1.96 \times \frac{2.89}{\sqrt{60}}$$
$$0.1\% \leq \bar{R} \leq 1.56\%$$

The constant 1.96 is obtained from the fact that $\Pr(z > 1.95) \approx 2.5\%$, where z is a standard normal random variable. Given the large estimation error of this strategy, the true annual return is estimated to be 95% likely to lie between 1.2% and 18.7%, a range found by multiplying the monthly returns by 12. It would be a mistake to use 0.83% (10% annualized) as the input, as if it were the true expected return. Another hedge fund strategy with 20 years of data and an estimated mean of 0.83% should be preferred to the new strategy due to its reduced estimation risk compared to funds with short return histories. How can estimation risk be incorporated into the mean-variance optimization problem?

Since estimation risk is an important consideration, dealing with it has a long history. Although a full discussion of its different treatments is beyond the scope of this book, robust optimization can be discussed as one particular approach. **Robust optimization** selects final solutions that incorporate estimation error directly into the modeling process. A simple version is presented here.

Suppose there are N assets with estimates of their expected returns, $\bar{R}_i$ for $i = 1, \ldots, N$. Similar to the previous example, suppose we have some ideas about the accuracy of our estimates. In particular, if $\bar{R}_i^*$ is the true value of the mean return, then a 95% confidence in the estimate can be expressed as:

$$-\varepsilon_i \leq (\bar{R}_i - \bar{R}_i^*) \leq \varepsilon_i \tag{2.8}$$

That is, the analyst can be 95% confident that the error in the estimate of the mean return does not exceed ε_i in absolute terms. If returns are normally distributed, then the estimation error term (ε_i) can be expressed as $1.96 \times \sigma_i/\sqrt{T}$ for a 95% confidence level, where σ_i is the standard deviation of the asset return, and T is the number of observations.

The process of incorporating estimation risk into the optimization is to select the optimal weights using estimated expected means that are reduced to the lower value of the selected confidence interval. Including the lower value of the confidence interval displayed in Equation 2.8 in the optimization problem, the problem will simplify into a regular mean-variance optimization by replacing $\bar{R}_i$, the estimated mean return of asset i, with $(\bar{R}_i - \varepsilon_i)$:

$$\bar{R}_i \rightarrow (\bar{R}_i - \varepsilon_i) \tag{2.9}$$

That is, the estimated expected mean returns are replaced according to Equation 2.9, and then the optimization proceeds in the normal way, as if there were no estimation risks.[7] It is important to notice the impact of the estimation risk on the optimal allocation. The estimated expected returns on assets are reduced by the size of the estimation error. That is, the larger the assumed estimation error, the lower the mean that is used to perform the optimization. Consequently, those asset classes for which the portfolio manager has the least estimation confidence will be most penalized and will receive lower allocations than would have been received if the estimation risk were ignored.

For example, suppose the mean return from a new strategy is estimated to be 1% per month. Assume that given the volatility of returns and the number of observations, the 95% confidence interval for the estimated mean is 1% ± 1.2%. The input to the optimization based on estimation risk is simple. Instead of using 1% as the expected mean return of the strategy, we should use 1% − 1.2% = −0.2%. That is, in the presence of estimation risk, the optimization constructs the optimal portfolio as if there were no estimation risk and with the adjusted expected rate of return on the asset being equal to −0.2%. Since short sales are ruled out, the optimal allocation to this strategy may turn out to be zero if it does not provide enough diversification benefits. It is important to stress that the expected returns on all assets will likely need to be reduced somewhat, because it is reasonable to assume that no expected return can be estimated without any error. So the net impact is not going to be limited to a reduction in allocation to some specific assets, but rather the entire allocation to risky assets may be reduced.

2.3 RISK BUDGETING

Risk budgeting refers to a broad spectrum of approaches to portfolio construction and maintenance that emphasize the selection of a targeted amount of risk and the allocation of that aggregate portfolio risk to various categories of risk. A risk budget is analogous to an ordinary budget for expenses in which an aggregate level of expenses is determined and is spent among various categories. Similarly, a risk budgeting process might decide that an appropriate level of risk is 15% and then might allocate that acceptable level of risk among the various asset classes.

2.3.1 Specifications in Risk Budgeting

Risk budgeting requires a specification of how risk is measured, but the risk budgeting approach can be used with virtually any quantitative approach.

Common examples of risk measurements to use in a risk budgeting process include standard deviation of returns, standard deviation of tracking error against a benchmark, value at risk, and beta. Risk budgeting requires a clear specification of the relationship between the measured risk of the total portfolio and the measured risks of the portfolio's constituent assets. The quantitative link between the risk of the portfolio and the risks of its constituent assets permits risk to be budgeted among the available assets.

Risk budgeting does not require the specification of expected returns. However, a popular application of risk budgeting allocates a portfolio between passive investments (e.g., indexation) and active investments (e.g., alternative investments) based in part on estimates of the extent to which the active investments can be expected to have higher expected returns than the passive investments. For example, a risk-budgeting framework may be designed to guide the asset allocator into deciding how much risk out of a total risk budget of 15% to allow for actively managed investments such as hedge funds in the pursuit of earning potential alpha. Other risk budgeting approaches may use a mean-variance framework in which the expected return of every asset is specified.

One case where expected returns are used in risk budgeting is when the asset allocator is using the standard deviation of the tracking errors between the portfolio's return and a benchmark's return. The asset allocator has to decide which managers will be allowed to deviate from a benchmark. One important factor influencing this decision is the potential alpha of the manager. For instance, an asset allocator may decide that a large-cap equity investor is not likely to generate substantial alpha; therefore, the portfolio manager will not be allocated any tracking error risk. In other words, an index fund will be used for this asset class. On the other hand, the asset allocator will be willing to spend a substantial portion of the portfolio's tracking error risk on an equity long/short manager who has the potential to generate a significant amount of alpha.

2.3.2 Implementing a Risk Budgeting Approach

The key to risk budgeting is its focus on risk allocation as the primary driver of the portfolio selection and monitoring process. Risk budgeting differs from mean-variance optimization, in which the optimizer selects assets driven by the trade-off between risk and return.

As previously mentioned, to implement and understand risk budgeting, the first step is to come up with a measure of a portfolio's total risk. The two most common measures are standard deviations of returns and value at risk. Note that under the assumption that returns are normally distributed, there is a one-to-one relationship between standard deviation and value of risk, and therefore either one could be used as a measure of total risk. In this section, standard deviation of returns will be used as the portfolio's measure of total risk.

The CAIA Level I book explains that the variance of a portfolio's return can be written as:

$$\sigma_P^2 = \text{Var}\left[\sum_{i=1}^{N} w_i R_i\right] = \sum_{i=1}^{N}\sum_{i=1}^{N} w_i w_j \sigma_{ij} \tag{2.10}$$

Here, w_i is the weight of asset i in the portfolio, and σ_{ij} is the covariance between asset i and asset j. Risk budgeting attempts to measure the contribution of each asset class to the total risk of the portfolio, typically measured by the standard deviation of returns, σ_P. Risk budgeting can be used to measure the risk contributions of asset classes as well as contributions of risk factors.

Having selected standard deviation to represent the total risk of a portfolio, measuring each asset class's contribution to the total risk is not complicated. In particular, the total risk can be decomposed in the following form:

$$\sigma_P = \frac{\partial \sigma_P}{\partial w_1} w_1 + \cdots + \frac{\partial \sigma_P}{\partial w_N} w_N \tag{2.11}$$

Equation 2.11 states that the total risk of a portfolio can be decomposed into N components, with each measuring the contribution of an asset class to the total risk. The contribution of each class is measured by the asset's weight in the portfolio multiplied by the sensitivity of the portfolio's standard deviation to small changes in the weight of the asset ($\partial \sigma_P / \partial w_i$). It can be seen that an asset that has a relatively large weight is likely to contribute a relatively large amount to the total risk of the portfolio. However, depending on the volatility of the asset and its correlations with other assets in the portfolio, the actual contribution of the asset could be much smaller or larger than its weight in the portfolio.

A simple analytical formula allows us to calculate the risk contribution of each asset class. The total risk of a portfolio can be decomposed into the contribution of each asset class to the total risk. Because the contribution of each asset class to the total risk was measured by $(\partial \sigma_P / \partial w_i) \times w_i$, we can use the formula for the standard deviation of the portfolio to evaluate this contribution:

$$\frac{\partial \sigma_P}{\partial w_i} \times w_i = \frac{\sigma_{iP}}{\sigma_P} \times w_i = \rho_i \times \sigma_i \times w_i \tag{2.12}$$

where σ_{iP} is the covariance of asset i with the portfolio, and ρ_i is the correlation of asset i with respect to the portfolio. According to Equation 2.12, the contribution of each asset depends on its correlation with the portfolio, its own volatility, and its share in the portfolio. An increase in each of these three variables will increase the risk contribution of the asset. Note that given the definition of beta, $\beta_i = \rho_i \times (\sigma_i / \sigma_P)$, the relationship expressed in Equation 2.12 can be stated in terms of betas as well.

2.3.3 A Three-Asset Example of Risk Budgeting

Consider the information in Exhibit 2.3 about a portfolio consisting of three assets. The diagonal terms of the variance-covariance matrix represent the variances of the assets. For example, the standard deviation of Asset 1 is $\sqrt{0.00780} = 8.83\%$.

Using the information provided in Exhibit 2.3, it can be seen that the mean and the standard deviation of the portfolio will be 9.20% and 8.11%, respectively. The risk decomposition of this portfolio will be

$$8.11\% = \underset{\text{Asset 1 Contribution}}{(7.26\% \times 40\%)} + \underset{\text{Asset 2 Contribution}}{(6.93\% \times 35\%)} + \underset{\text{Asset 3 Contribution}}{(11.12\% \times 25\%)}$$

EXHIBIT 2.3 Properties of Three Hypothetical Assets

			Variance-Covariance Matrix		
	Mean	**Weights**	**Asset 1**	**Asset 2**	**Asset 3**
Asset 1	10%	40%	0.00780	0.00746	0.00064
Asset 2	12%	35%	0.00746	0.01020	–0.00372
Asset 3	4%	25%	0.00064	–0.00372	0.04025

For example, the contribution of Asset 2 to the total risk of the portfolio is 6.93% × 35% = 2.43%. These figures were calculated with the help of Equation 2.12. Let's consider another example to see how the equation can be applied.

Suppose the weights of three asset classes in the portfolio displayed in Exhibit 2.3 are changed to 50%, 40%, and 10% for Assets 1, 2, and 3, respectively. What will be the risk contribution of Asset 1? First, we need to calculate the standard deviation of the portfolio:

$$\sigma_P = \sqrt{\sum_{i=1}^{3}\sum_{i=1}^{3} w_i w_j \sigma_{ij}} = 8.21\%$$

Next, the contribution of Asset 1 can be calculated using Equation 2.12 and the information provided by Exhibit 2.3. The correlation between Asset 1 and the portfolio is 0.9586, and its standard deviation was reported in Exhibit 2.3 to be 8.83%.[8]

$$\text{Contribution of Asset 1} = 0.9586 \times 0.0883 \times 50\% = 4.23\%$$

It can be seen that while the weight of Asset 1 in the portfolio is 50%, close to 52% (4.32%/8.21%) of the total risk comes from Asset 1.

Example: The correlation of Asset 3 with the previous portfolio is 0.1734 and its standard deviation is 20%. What is the contribution of Asset 3 to the total risk of the portfolio?

$$\text{Contribution of Asset 3} = 0.1734 \times 0.2 \times 0.1 = 0.35\%$$

In this case, Asset 3 is 10% of the portfolio, but it contributes only 0.35% to the total risk, or only 4.2% (0.35%/8.21%) of the total risk is due to Asset 3.

To summarize, the total risk of a portfolio can be decomposed so that the contribution of each asset class to total risk is measured. This decomposition increases the portfolio manager's understanding of how the portfolio is likely to react to changes in each asset class. In addition, once the risk budget of each asset class is measured, the portfolio manager may consider changing the allocation so that each asset class's contribution does not exceed some predetermined risk budget. Finally, the mean-variance optimization problem discussed earlier in this chapter can be adjusted to incorporate constraints related to the risk budget.

2.3.4 Applying Risk Budgeting Using Factors

The preceding discussion focused on risk budgets associated with asset classes. The total risk of a portfolio was decomposed into the contribution of each asset class to that total risk. It is possible to use the same approach to decompose the total risk of a portfolio by measuring the contribution of each risk factor to the total risk. A number of macroeconomic and financial factors can affect the performance of a portfolio. An asset owner may already have significant exposures to some of these factors. For example, the manufacturer of a product that is sold in foreign markets may have significant exposures to currency risk. That may lead the asset owner to instruct the portfolio manager to measure and then limit the exposure of the portfolio to currency risk. On the other hand, the same asset owner may wish to measure and adjust the risk exposure of the portfolio to the interest rate factor because it needs the assets to fund a liability that is interest rate sensitive. Here we discuss how the contribution of factor volatilities to total risk of a portfolio can be measured.

To see this, suppose there are two risk factors, F_1 and F_2, that are the major drivers of the portfolio's return. Their degree of importance can be measured by regressing the portfolio's rate of return on these two factors (e.g., one factor could be changes in the credit spread and the other could be changes in the price of oil).

$$R_{Pt} = a + b_1F_{1t} + b_2F_{2t} + \varepsilon_t \tag{2.13}$$

Here, R_{Pt} is the rate of return on the portfolio; a, b_1, and b_2 are the estimated parameters of the regression model; and ε_t is the residual part of the regression that represents the part of the return that cannot be explained by the two factors. The total risk of the portfolio can now be decomposed into the contributions of each risk factor:

$$\sigma_P = \underbrace{\left(\rho_{F_1} \times \sigma_{F_1} \times b_1\right)}_{\text{Contribution of Factor 1}} + \underbrace{\left(\rho_{F_2} \times \sigma_{F_2} \times b_2\right)}_{\text{Contribution of Factor 2}} + \underbrace{\left(\rho_{\varepsilon} \times \sigma_{\varepsilon}\right)}_{\text{Contribution of Unknown Sources}} \tag{2.14}$$

Here, ρ_{F_1}, ρ_{F_2}, and ρ_ε are the correlations of the two factors and the residual risk with the portfolio's return, respectively. Each of the first two terms that appear on the right-hand side of Equation 2.14 represents the contribution of a factor to the total risk of the portfolio. The last term represents the contribution of unknown sources of risk.

For example, suppose the correlation between changes in the oil price and the return on the portfolio displayed in Exhibit 2.3 is 0.31. The standard deviation of changes in the oil price is estimated to be 20%, and the factor loading of the portfolio on oil (i.e., the coefficient of oil in Equation 2.13) is 0.757. What is the risk contribution of oil to the total risk of the portfolio?

$$\text{Contribution of Oil} = 0.31 \times 0.20 \times 0.757 = 4.69\%$$

Therefore, 4.69% of the total risk of 8.11% of the portfolio described in Exhibit 2.3 can be contributed to the volatility in oil prices. It can be seen how the risk budget associated with risk factors can be helpful in understanding the risk profile of a portfolio. In addition, limits on risk budgets associated with risk factors can be incorporated into the mean-variance optimization. In fact, we saw a version of this earlier in the chapter when we discussed how limits on factor exposures can be added to the mean-variance model.

So far, the measure of total risk has focused on standard deviation. Value at risk (VaR) and conditional value at risk (CVaR) can also be used to measure total risk. Much of the discussion presented here can be represented using these two measures of total risk with minimal change. For further discussion of these measures of total risk and risk budgeting, see Pearson (2002).

2.4 RISK PARITY

While risk budgeting is mostly concerned with measuring the exposure of a portfolio, the **risk parity** approach uses risk budgeting results and attempts to create portfolios that equally weight the risk contributions of each asset or asset class in a portfolio. In risk parity, a portfolio allocation model is constructed based entirely on the risk contribution of each asset to the total risk of the portfolio, with no consideration of expected return of each asset class. Specifically, and as is detailed later, the risk-parity approach recommends that the allocation to each asset class should be set so that each asset class has the same marginal contribution to the total risk of the portfolio. The result is that the allocation to each asset tends to be inversely proportional to its risk once the diversification effect is taken into account. In risk parity, if equity is viewed as an asset that contributes more risk to the portfolio than bonds do, then less of the portfolio should be allocated to equities than to bonds.

2.4.1 Three Steps in Implementing the Risk Parity Approach

There are three steps in implementing the risk parity approach:

First, similar to risk budgeting, risk parity requires a definition of the total risk of a portfolio. Risk parity does not impose a uniform measure of total risk. However, total risk is typically measured by the standard deviation of the rate of return on the portfolio. Alternatively, one could use value at risk (VaR) or some other measure of risk. The advantage of using VaR as a measure of total risk is that one can incorporate skewness and kurtosis into the measure of total risk. For the purpose of this discussion, standard deviation is used as a measure of total risk.

Second, risk parity requires a method to measure the marginal risk contribution of each asset class to the total risk of the portfolio. The marginal risk contribution of an asset to the total risk of a portfolio indicates the rate at which an additional unit of that asset would cause the portfolio's total risk to rise. The marginal risk contribution of an asset depends on the composition of the portfolio. For example, adding a hedge fund to an otherwise diversified portfolio may contribute little or no risk, since the hedge fund may offer substantial diversification benefits. However, as the hedge fund's allocation to the portfolio is increased, the effect of additional allocations of hedge funds (i.e., the marginal contribution) also increases. Accordingly, at high levels of allocation to hedge funds, additional allocations may increase risk substantially, as the portfolio becomes concentrated in hedge funds rather than diversified. The measurement of risk contributions was discussed in the previous section.

Finally, portfolio weights are determined for all available assets. The weights are typically computed using a trial-and-error process until the marginal contributions from all assets to the total risk of the portfolio are equal. The previous section showed that the marginal contribution of an asset class to the total risk of a portfolio is given by:

$$\text{Contribution of Asset } i = \frac{\partial(\text{Total Risk})}{\partial(\text{Weight of Asset } i)} \times (\text{Weight of Asset } i) \qquad (2.15)$$

Therefore, the risk parity approach to asset allocation seeks a portfolio in which all asset classes (except cash) contribute the same amount to the total risk of the portfolio. It is important to note that there is no volatility target, and the total risk of the portfolio is endogenous to the process. That is, the weights are numerically adjusted to create a portfolio in which each asset contributes the same amount to the total risk.

2.4.2 Creating a Portfolio Using the Risk Parity Approach

This section addresses the central point of the risk parity approach: how to determine portfolio weights. Equations 2.11 and 2.15 demonstrate that, in all cases, the total risk of a portfolio may be expressed as the sum of the marginal contributions of the portfolio's constituent assets. The risk parity approach is the simple prescription that the portfolio's weights should be selected such that the marginal contribution of each asset is equal. Thus, to create a portfolio of N assets using the risk parity approach, the weights need to be adjusted until the marginal contribution to risk for each asset in the portfolio is equal to $(1/N)$ times the total risk of the portfolio. The portfolio weights that equalize all the marginal contributions to risk can be easily found using a trial-and-error approach or an optimization package such as Microsoft Excel's Solver.

Consider the information for three asset classes in Exhibit 2.4.

The data from Exhibit 2.4 can be used to generate portfolio weights using the risk parity approach. A trial-and-error search can lead to the risk parity solutions depicted in Exhibit 2.5. For comparison purposes, the weights associated with other approaches are presented as well.

EXHIBIT 2.4 Properties of Three Asset Classes

			Correlation Matrix		
January 1990– December 2015	Monthly Mean	Monthly Standard Deviation	MSCI World	Barclays Global Aggregate	HFRI Fund Weighted Composite
MSCI World	0.66%	4.42%	1.000	0.302	0.699
Barclays Global Aggregate	0.49%	1.56%	0.302	1.000	0.156
HFRI Fund Weighted Composite	0.83%	1.95%	0.699	0.156	1.000

Source: Bloomberg, HFRI, and authors' calculations.

EXHIBIT 2.5 Portfolio Weights and Their Properties

Portfolio Weights			
January 1990–December 2015	**(55/35/10) Portfolio**	**Risk Parity Portfolio**	**Minimum Volatility Portfolio**
MSCI World	55.00%	14.33%	0.00%
Barclays Global Bond Aggregate	35.00%	50.59%	63.04%
HFRI Fund Weighted Composite	10.00%	35.09%	36.96%
Performance			
Monthly Mean	0.62%	0.63%	0.62%
Monthly Standard Deviation	2.78%	1.60%	1.30%

Source: Bloomberg, HFRI, and authors' calculations.

There are three different portfolios in Exhibit 2.5. The first one is constructed using no optimization or risk parity. The risk parity portfolio is constructed to equalize the risk contributions of the three asset classes. The minimum volatility portfolio is constructed using mean-variance optimization, in which the goal is to use positive weights to create a portfolio with minimum standard deviation regardless of the mean. It can be seen that the risk parity portfolio allocates relatively high weights to bonds and hedge funds. The minimum volatility portfolio has no allocation to equities. The risk contributions of the three asset classes for each of the three portfolios are presented in Exhibit 2.6.

As expected, in the risk parity portfolio, each asset contributes the same marginal risk (0.53%), which is 33.3% of the resulting portfolio's total risk. The process is iterative, because the total risk of the portfolio changes as the allocations are changed.

Note that in this example risk parity requires a substantial allocation to fixed income. This is because the fixed-income investment exhibits lower total risk; therefore, more of the portfolio can be allocated to fixed income while keeping its marginal contribution to the portfolio's total risk the same as the marginal contributions of equities and hedge funds. The risk parity approach typically prescribes a low-risk portfolio by overweighting low-risk assets relative to the market portfolio.

EXHIBIT 2.6 Risk Contributions of the Three Asset Classes

January 1990–December 2015	**(55/35/10) Portfolio**	**Risk Parity Portfolio**	**Minimum Volatility Portfolio**
MSCI World	2.39%	0.53%	0.00%
Barclays Global Aggregate	0.26%	0.53%	0.82%
HFRI Fund Weighted Composite	0.14%	0.53%	0.48%
Total*	2.78%	1.60%	1.30%

* Because of rounding error the columns do not add up to the total.
Source: Bloomberg, HFRI, and authors' calculations.

2.4.3 The Primary Economic Rationale for the Risk Parity Approach

The portfolio in Exhibit 2.5 would have performed very well over the past 20 years because the portfolio was allocated heavily to fixed-income assets, and fixed-income assets outperformed equity assets on a risk-adjusted basis during that period. However, selecting strategies based solely on successful historical performance, even spanning 20 years, runs the risk of unsuccessfully chasing historical performance. Chasing historically superior risk-adjusted performance is futile in an efficient market.

In theory, it is difficult to find any reason why a risk parity portfolio should be optimal. For example, if asset returns are normally distributed and financial markets are perfect, then there are sound economic reasons why investors should select portfolios that plot on the efficient frontier. In other words, it would be difficult to come up with sound economic reasons supporting the risk parity approach in perfect financial markets. In addition, it is not immediately clear how and why market imperfections should make risk parity portfolios more desirable compared to, say, mean-variance efficient or even equally weighted portfolios. Therefore, the case for the risk parity approach must be built through careful analysis of its properties.

As was previously shown and as demonstrated by other studies, the risk parity approach creates low volatility portfolios where low-risk assets are overweighted. Under what conditions, then, would low-risk portfolios become optimal for some investors? One obvious answer is that if an investor has a very high degree of risk aversion, then a risk parity portfolio could be optimal for him. However, a low volatility portfolio can also be constructed using the mean-variance framework. For example, one can create a minimum volatility portfolio using the mean-variance approach (see Exhibits 2.5 and 2.6). It turns out that the most compelling arguments that can be put forward in support of the risk parity approach also support the use of low-risk portfolios, including the minimum volatility approach.

The economic rationale for low volatility portfolios is that because of market imperfections, many investors are unable or unwilling to use leverage. This is referred to as leverage aversion. The **leverage aversion theory** argues that large classes of investors cannot lever up low volatility portfolios to generate attractive returns and that, as a result, low volatility stocks and portfolios are underpriced. While leverage increases risk and excess return by the same factor, under some conditions it might be possible to create portfolios with higher Sharpe ratios by applying leverage to low-risk portfolios. As an example, consider the two portfolios depicted in Exhibit 2.7.

While the expected rate of return on the RP portfolio may appear unattractively low to some investors, those who are willing and able to acquire funding at

EXHIBIT 2.7 Two Hypothetical Portfolios

	Expected Return	Standard Deviation	Sharpe Ratio
Risk Parity Portfolio	8.00%	10.00%	0.60
Minimum Volatility Portfolio	11.00%	18.00%	0.50
Riskless Rate	2.00%		

attractive rates can lever up the RP portfolio and earn a superior risk-adjusted return compared to the MV portfolio. In particular, a portfolio with 60% leverage will produce (11.6% = 1.6 * 8% – 0.6 * 2%) return and (16% = 1.6 * 10%) volatility.

The important question is why the low volatility portfolio should provide a higher Sharpe ratio in the first place. The leverage aversion argument is that because many investors are not allowed to use leverage, demand for low-risk stocks is low, and therefore they are undervalued. For example, if a mutual fund manager wants to overweight stocks he judges to be undervalued but still track a broad equity benchmark with some degree of accuracy, the manager cannot afford to overweight low volatility stocks. Although low-risk stocks may offer attractive risk-adjusted returns, the mutual fund manager will not be able to lever up their low raw returns to match the overall market return; as a result, the fund will underperform its benchmark in terms of raw returns. The idea that low volatility stocks are underpriced and therefore offer higher expected risk-adjusted returns is known as the **volatility anomaly**.[9] According to this anomaly, portfolios consisting of low volatility stocks have historically outperformed the overall market. However, the evidence seems to indicate that the anomaly has weakened since its discovery by academic researchers. Related to the volatility anomaly is the **betting against beta anomaly**, which has documented that portfolios consisting of low-beta stocks have outperformed the market in the past.[10] The explanation for the betting against beta anomaly is rather similar to the one set forth for the low volatility anomaly. Many investors are unwilling or unable to use leverage to increase the betas of low-beta portfolios. During periods of rising market prices, investors want to be invested in high-beta stocks. As a result, high-beta stocks are bid up, which reduces their future returns. Therefore, those investors who are willing and able to use leverage should experience a higher risk-adjusted return through investing in low-beta stocks.

2.4.4 Four Rationales for Risk Parity That Can Be Rejected

Market participants have set forth other arguments in support of the risk parity approach, such as that it has performed well in the past, that it produces well-balanced portfolios, or that the resulting portfolios are well diversified in terms of risk. These arguments can be rejected. First, it is clear that going forward, low-risk fixed-income instruments of developed and most emerging economies will not perform as well as they did during the past 20 years, as many bonds now have yields far below the historical returns in their respective markets. Therefore, risk parity portfolios that overweight fixed income are unlikely to repeat their historical performance.

Second, the argument that the portfolios will be well balanced is imprecise and appears to belong to marketing materials rather than an economically sound investment report. It is unclear what is meant by "well balanced" and why it should lead to superior risk-adjusted performance.

Third, it is important to note that risk parity may introduce risks that are absent in other strategies. The fact that risk parity portfolios require investors to use leverage may indicate that they are not directly comparable to equally volatile portfolios that do not use leverage. Funding liquidity risk, which was discussed earlier in this chapter, introduces a risk associated with leverage that is not present in unlevered portfolios. For example, during periods of market stress, the investor may be asked to reduce the portfolio's leverage and therefore liquidate the portfolio at the most inopportune time.

A fourth unsupported rationale for risk parity is that it can exploit anomalies that exist in alternative asset markets. However, there have been no studies to show that the low volatility or betting against beta anomalies work in the alternative investment area. That is, the low-risk strategies may or may not provide the highest risk-adjusted returns. While there is some evidence that levered low-risk portfolios of traditional asset classes may provide attractive risk-adjusted returns, there is no evidence that such a strategy could work in the alternative investment area. Compared to portfolios of liquid traditional assets, funding of levered portfolios of alternative investments is likely to be more difficult and more expensive.

It is important to note that because some alternative investments have low volatility and low correlations with other asset classes, the allocations to alternative investments using the risk parity approach will be relatively high compared to market weights and typical institutional portfolios. Exhibit 2.5 is an example of how risk parity may lead to unusually high allocations to low-risk investments, including alternatives. There is an equally high allocation to the HFRI index in the minimum volatility portfolio. As previously stated, risk parity relies on leverage aversion and low volatility anomalies to justify increased allocations to low-risk assets. However, using leverage to increase the return on a low volatility portfolio that contains significant allocations to alternative assets may not be possible or desirable.

While risk parity may be a viable approach to asset allocation, it does not represent a trading strategy that can be employed by active managers seeking to maximize risk-adjusted return, because risk parity does not require or use any estimate of expected return. The risk parity approach may be suitable for institutional and high-net-worth investors who do not face substantial constraints on their asset allocation policies and who are able to use leverage to adjust the total risk to meet their target total risk.

2.4.5 Equally Weighted and Volatility-Weighted Portfolios

Risk parity is one approach to creating a low volatility portfolio. Another approach is to use an equally weighted portfolio. An equally weighted portfolio is, by definition, rather well diversified and, in practice, is likely to have relatively high allocations to less risky assets. The reason is that in most applications, equity serves as both the largest asset class and the riskiest. Thus, equal weighting typically underweights equities relative to a market portfolio. Another approach, as seen previously, is to use mean-variance optimization to identify the minimum variance portfolio.

Finally, a volatility-weighted portfolio can be used to create a low volatility portfolio by weighting each asset inversely to its volatility. In this approach, the weight of each asset class is shown in Equation 2.16.

$$w_i = \frac{\sigma_i^{-1}}{\sum_{j=1}^{N} \sigma_j^{-1}} \qquad i = 1, \ldots, N \tag{2.16}$$

EXHIBIT 2.8 Portfolio Weights and Their Properties

		Portfolio Weights		
January 1990–December 2015	**(55/35/10) Portfolio**	**Risk Parity Portfolio**	**Volatility-Weighted Portfolio**	**Minimum Volatility Portfolio**
MSCI World	55.00%	14.33%	16.37%	0.00%
Barclays Global Aggregate	35.00%	50.59%	46.49%	63.04%
HFRI Fund Weighted Composite	10.00%	35.09%	37.13%	36.96%
		Performance		
Monthly Mean	0.62%	0.63%	0.64%	0.62%
Monthly Standard Deviation	2.78%	1.60%	1.67%	1.30%

Source: Bloomberg, HFRI, authors' calculations.

This means that the portfolio weight for each asset class is proportional to the inverse of its volatility. The denominator is the sum of the inverses of the return volatilities of all assets. This ensures that the weights will add up to one.

The volatility-weighed approach and the risk parity approach are rather similar. The difference between the two is that the risk parity approach takes into account the diversification that each asset offers, whereas the volatility-weighted approach allocates the portfolio solely on the basis of each asset's stand-alone risk (i.e., each asset's volatility). The volatility-weighted approach is identical to the risk parity approach when there are only two assets or when all correlations between each pair of assets are the same. In other words, if all correlations are identical, all assets offer equal diversification benefits, and thus risk parity and volatility weighting are equivalent. Exhibits 2.8 and 2.9 expand the results reported in Exhibits 2.5 and 2.6 to include information on a volatility-weighted portfolio.

One advantage of the risk parity, volatility-weighted, and minimum volatility approaches displayed in Exhibits 2.8 and 2.9 is that they do not require estimates of expected returns as inputs. As discussed in Chapter 1, expected returns are very

EXHIBIT 2.9 Risk Contributions of the Three Asset Classes

January 1990–December 2015	**(55/35/10) Portfolio**	**Risk Parity Portfolio**	**Volatility-Weighted Portfolio**	**Minimum Volatility Portfolio**
MSCI World	2.39%	0.53%	0.63%	0.00%
Barclays Global Aggregate	0.26%	0.53%	0.46%	0.82%
HFRI Fund Weighted Composite	0.14%	0.53%	0.58%	0.48%
Total*	2.78%	1.60%	1.67%	1.30%

* Because of rounding errors, the columns do not add up to the total.
Source: Bloomberg, HFRI, authors' calculations.

difficult to predict, and a long history of returns is needed to obtain accurate estimates. This is particularly relevant when alternative investments are considered, because many of them lack the long history that traditional asset classes have.

2.5 FACTOR INVESTING

Factor investing is a recent development in the area of asset allocation, and some aspects of it are closely tied to the developments of certain hedge fund strategies. This section offers a brief introduction to this topic; detailed discussions of this approach can be found in Ang (2014). The basic ideas behind factor investing were developed by academic researchers over the past 30 years. However, only in recent years have these ideas been synthesized and presented in practical form by the investment industry.

2.5.1 The Emergence of Risk Factor Analysis and Three Important Observations

One of the central concepts in finance is that returns above the risk-free rate are compensation for exposure to risks. Stocks have earned a premium above the risk-free asset because they expose their owners to their underlying risk factors. The capital asset pricing model (CAPM) can be thought of as the first theory of risk factors. According to the CAPM, the expected excess return on a risky asset is equal to the beta of the asset times the risk premium on the market portfolio. When asset returns are normally distributed, the CAPM correctly identifies the only risk factor to be the return on the market portfolio, and the measure of exposure of each stock to this factor to be its beta. By relaxing the assumptions of the CAPM, academic and industry researchers have extended the model, leading them to develop a long list of risk factors.[11] Even though there is no consensus about the number of risk factors in financial markets, it is well established that the market portfolio is not the only risk factor in the market. The risk premiums earned by stocks and other traditional assets are indeed a combination of various risk premiums.

Our discussion of factor investing begins with three important observations detailed by Ang (2014):

1. Factors matter, not assets: The factors behind the assets matter, not the assets themselves. Assets are nothing more than means for accessing factors.
2. Assets are bundles of factors: Most asset classes expose investors to a set of risk factors; therefore, the risk premium offered by an asset class represents a package of risk premiums offered by factor exposures. Exposures of assets and risk premiums to factors can be used to estimate the risk premium that an asset should provide.
3. Different investors should focus on different risk factors: Asset owners differ in terms of time horizons, risk tolerances, objectives, and liabilities that should be funded. These differences require careful examination of factor exposures of a portfolio in order to ensure that the asset owners do not have too much or too little exposure to some factors.

If we consider the risk premium earned by each asset class to be a function of exposures to several risk factors, then several important questions arise:

- What is a risk factor?
- Do all risk factors offer the same risk premium?
- How do these risk premiums behave through various stages of a business cycle?
- Are all risk factors investable?
- How does one perform risk allocation based on risk factors?
- Do allocations based on risk factors outperform allocations based on asset classes?

These questions are addressed in the next six sections.

2.5.2 How Risk Factors Are Described

What is a risk factor? A risk factor represents a unique source of risk and risk premium in financial markets such that the observed risk and risk premium cannot be fully explained by other risk factors. In other words, risk factors are not supposed to be highly correlated with each other. In addition, risk factors should have a sound economic foundation, with rigorous academic and industry research supporting their presence. Finally, risk factors must show that their risk premiums are persistent over long periods.

Let us examine two well-known equity risk factors: momentum and value factors. To create the momentum factor, researchers examine the performance of equity prices through time. Then two portfolios are created at the end of each period (e.g., month). One equally weighted portfolio will contain those stocks that performed well during some specific period (e.g., past 12 months), and the other equally weighted portfolio will contain those stocks that performed poorly during the same period. This procedure is repeated every period and results in two return series. If one were to go long the portfolio of winners and go short the portfolio of losers, the return to this active strategy would be the return to the momentum factor. We can ascertain that this is indeed a new factor if the return to this strategy cannot be explained by other factors, such as return to the market portfolio.

The value factor is constructed the same way except that stocks are sorted based on the ratio of their book values to their market values. Stocks with high ratios of book value to market value are considered value stocks, and those will low ratios are considered growth stocks. Two equally weighted portfolios are created each period. One will consist of those stocks with above-average book-to-market ratios, and the other one will consist of those stocks with below-average book-to-market ratios. Again, a strategy will go long the first portfolio and short the second portfolio. The return to this active strategy will represent the return to the value factor. Exhibit 2.10 displays properties of these two risk factors.

It is important to note that the average returns reported in Exhibit 2.10 are for portfolios that, at least in theory, do not require any investment. This is because each factor consists of equal-size long and short positions. Therefore, the initial capital is available to be invested in the riskless assets, the market portfolio, or some combination of the two. For instance, if the capital were invested in the market, then the entire portfolio would earn the market portfolio plus the return on each factor. Of

EXHIBIT 2.10 Properties of Value and Momentum Factors

1927–2015	Value	Momentum	50/50 Portfolio
Annualized Mean	4.58%	8.23%	6.41%
Annualized Standard Deviation	12.25%	16.43%	8.01%
Correlation with Market Return	0.23	–0.34	–0.17
Correlation of Two Factors	–0.41		

Source: K. French Data Library.

course, the resulting portfolio would be highly volatile because it would have a gross leverage of 300%. Also, note that during the same period, the return on the market in excess of the riskless rate and its standard deviation were 7.72% and 18.71%, respectively.

Several important observations can be made from Exhibit 2.10. First, both risk factors have generated significant positive returns over a long period. Second, both risk factors are volatile, which means they may not consistently generate positive returns, and there will be periods during which they could generate negative returns. Third, the two factors are negatively correlated (–0.41) with each other and are not highly correlated with the market risk. The fact that the two factors have been negatively correlated in the past is one of the key benefits of factor investing in the sense that two sources of returns have been discovered that are negatively correlated. This means a portfolio of the two should be far less volatile, which is confirmed by the last column of the exhibit. From Exhibit 2.10 the potential benefits of factor investing can already be seen. By isolating risk factors and investing in them, one might be able to avoid those risks that are not rewarded by the marketplace and, as a result, create portfolios that display more attractive risk-return properties.

As was mentioned, academic and industry research has uncovered a number of risk premiums. Following is a partial list of the best-known risk factors:

- Value premium: Based on the return earned by stocks with high book values relative to their market values. The strategy is to take long positions in stocks with a relatively high ratio and short positions in stocks with a relatively low ratio.
- Size premium: Based on the rate of return earned by small-cap stocks. The strategy is to take long positions in small-cap stocks and short positions in large-cap stocks.
- Momentum premium: Based on past performance of stocks. The strategy is to take long positions in past winners and short positions in past losers.
- liquidity premium: Based on the risk premium earned by illiquid assets. The strategy is to take long positions in illiquid assets and short positions in similar but otherwise liquid assets.
- Credit risk premium: Based on the credit risk of bonds. The strategy is to take long positions in bonds with low credit quality and short positions in bonds with high credit quality.
- Term premium: Based on the riskiness of long-term bonds. The strategy is to take long positions in long-term bonds and short positions in short-term bonds.

- Implied volatility premium: Based on the risk premium earned by the volatility factor. The strategy is to be short the implied volatility of options (e.g., create a market-neutral position using short positions in out-of-the-money puts and short positions in stocks).
- Low volatility premium: Based on the return to low volatility stocks. The strategy is to take long positions in low volatility stocks and short positions in high volatility stocks. The same strategy can be implemented using betas.
- Carry trade: Based on the return to investments in high-interest-rate currencies. The strategy is to take long positions in bonds denominated in currencies with high interest rates and short positions in bonds denominated in currencies with low interest rates.
- Roll premium: Based on the return earned on commodities that are in backwardation. The strategy is to take long positions in commodities that are in backwardation and short positions in commodities that are in contango.

2.5.3 Risk Premiums Vary across Risk Factors

Do all risk factors offer the same risk premium? The answer to this question has already been provided by the two factors that were presented in Exhibit 2.10. Those two factors offered different risk-return profiles. Different risk factors offer different risk premiums, and the sizes of the risk premiums are not constant through time. Therefore, while a passive allocation to several risk factors could produce attractive results, a more sophisticated approach would consider the size of the premium attached to each risk factor and create an asset allocation that would take advantage of changes in risk premiums. In other words, it might be beneficial to apply tactical asset allocation to risk factors by assigning higher weights to risk factors that are believed to be offering more attractive risk premiums. In addition, academic research has shown that those risk factors that provide poor returns during bad times are the ones that provide attractive returns during normal times. Risk premiums associated with legitimate risk factors are there because these factors perform poorly during bad times. Investors should be willing to hold them only if they provide attractive returns during normal times. Actually, the first step in determining whether an observed source of return is a legitimate risk factor is to compare its return during good and bad times. If the factor provides an attractive return during good and bad times, then it is not a risk factor—it is an arbitrage opportunity.

2.5.4 Risk Factor Returns Vary across Market Conditions

How do these risk premiums behave through various stages of a business cycle? Exhibit 2.11 displays the rolling 24-month average returns on the two factors discussed in Exhibit 2.10.

Clearly, these two factors display significant time variations. As previously mentioned, we must expect factor premiums to display some volatility and to perform differently during various stages of the business cycle. If there were no risks, then there would be no premiums. For instance, the book-to-market factor (i.e., value factor) tends to produce attractive risk-adjusted returns during normal market conditions but have poor risk-adjusted returns during economic downturns.[12] The momentum

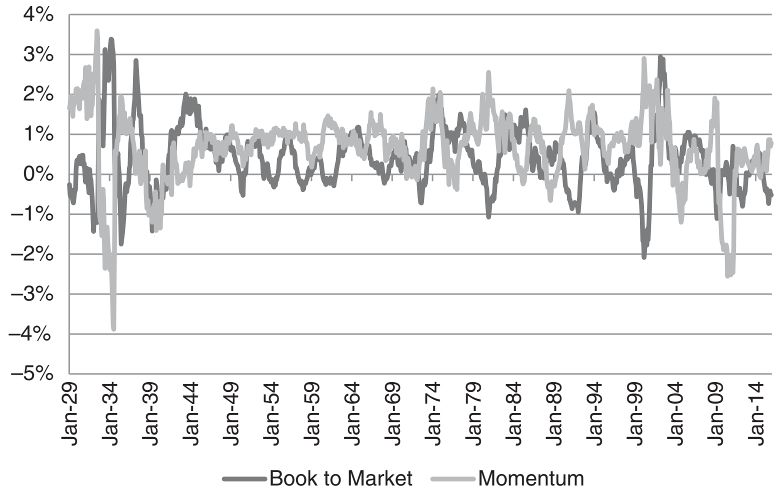


EXHIBIT 2.11 Rolling 24-Month Average Returns to Factors
Source: K. French Data Library.

factor tends to produce positive returns most of the time but has a tendency to display large negative returns over a short period of time when a market correction takes place. Momentum is notorious for performing poorly during momentum crashes. A **momentum crash** occurs when those assets with recent overperformance (i.e., those assets with momentum) experience extremely poor performance relative to other assets (Asness et al. 2010). In the same way that investors have learned to hold diversified portfolios in terms of asset classes, they should hold diversified portfolios in terms of risk factors.

2.5.5 Risk Factors and Investability

Are all risk factors investable? While 20 years ago it might have been difficult and costly to create investment strategies that represented various risk factors, financial innovations of the past two decades have substantially reduced the cost of such strategies. For example, in recent years, exchange-traded funds (ETFs) have become available that track value, volatility, momentum, size, and roll factors. These strategies are often described as using smart beta. Another term for describing the investability of a risk factor is that it is tradable.

While many factors are tradable, they may not be fully implementable. For example, the value factor has been shown to be strongest for the very small-cap firms. However, there is limited capacity for investing in these firms, and the bid-ask spreads for their stocks are quite wide. Since creating the value factor would require frequent rebalancing of the portfolio, it may not be possible to fully implement factor investing in value stocks. Therefore, traded investment products designed to replicate the value factor tend to concentrate their allocations to large-cap stocks, which have shown to be poor representatives of the value factor.

Even when pure risk factors may not be fully traded, their return properties are useful in measuring risk exposures of investment products (e.g., hedge funds) and in determining if they offer any alpha. If one were to regress the excess return of an investment product against the returns of several traded factors, the intercept would represent the alpha of the investment product.

2.5.6 Risk Allocation Based on Risk Factors

How does one perform risk allocation based on risk factors? Factor or risk allocation is no different from asset allocation. After all, one has to use assets to isolate risk factors. Two issues related to factor investing should be considered. First, it is not possible to create a benchmark for a passive factor investing strategy. This is because factors involve long/short strategies, and therefore it is unclear what the passive weights of a diversified portfolio of factors should be. Should the factors be equally weighted, or should they be volatility weighted? There is no clear answer to this question.

Second, direct factor investing requires the investor to actively manage portfolios that are supposed to represent risk factors. Therefore, there is no such thing as passively managed factor portfolios. It should be pointed out that if there are enough investment products (e.g., funds and ETFs) replicating each risk factor, then the investor could follow a buy-and-hold strategy involving these investment products. However, unless one is willing to allow the weights of some factor to become very high or low, the investor will have to rebalance the portfolio on a regular basis. In addition, direct investments may not be a viable strategy for some institutional investors, as most factor strategies would require long and short positions in certain asset classes, and many institutional investors may not be prepared to take short positions.

Several hedge fund strategies earn their returns by exploiting these risk factors. For instance, the merger arbitrage strategy earns a premium by exploiting a risk factor related to the uncertainty surrounding the completion of mergers. The convertible arbitrage strategy exploits a form of the implied volatility factor, and many equity long/short and equity market-neutral strategies have been shown to have significant exposures to risk factors of the equity markets (Zhang and Kazemi 2015). Finally, global macro strategies often rely heavily on the carry trade factor to generate returns.

2.5.7 Performance with Allocations Based on Risk Factors

Do allocations based on risk factors outperform allocations based on asset classes? In theory, allocations based on risk factors should provide the same risk-adjusted return as allocations based on assets if the same information sets are used (Idzorek and Kowara 2013). It can be shown that in a perfect world, where all risk factors can be traded and all asset returns can be explained by risk factors, neither approach is inherently superior to the other. Research using real-world data has demonstrated that either approach may be superior over a given period of time (Idzorek and Kowara 2013). In particular, in some cases, the apparent superiority of risk factors is a simple result of the fact that risk factors can be conceptualized as being an alternative set of asset classes that can be identified when the long-only constraint is removed. For instance, a portion of the return earned by the value factor is due to the fact that it shorts growth stocks. Therefore, an asset allocation strategy that would permit the investor to short stocks should be able to match the performance of a portfolio that allocates to the value factor. If all risk factors were traded and short sale constraints were removed, then it would be difficult to imagine that a portfolio constructed using risk allocation could outperform a portfolio constructed using asset allocation on a consistent basis.

Several practical issues associated with risk-factor-based asset allocation must be highlighted (Idzorek and Kowara 2013). First, portfolio construction using risk factors is not likely to become globally adopted, because it implies allocation strategies that are not sustainable (i.e., not consistent on the macro level). That is, not everyone can be short growth stocks or commodities that are in contango. Therefore, the capacity is likely to be limited, and as more money is allocated to factor investing, the strategies will become expensive and the risk premium will shrink or disappear altogether. Second, risk allocation requires asset owners to take extreme positions in some asset classes. Many institutional investors are not allowed to make such allocations. For example, taking short positions in all growth stocks, including such names as Google, Amazon, and Facebook, is not something that most investors are prepared to do. Third, risk allocation is not a magic bullet that will automatically lead to asset allocations that will dominate those based on asset classes. Similar to other investment products, the cost of the strategy must be taken into account. Flows into some of these factors have already reduced the size of the premium (e.g., size and low volatility factors have not performed well in recent years). Finally, alternative investments are important vehicles for accessing some of these risk factors, but they typically represent a bundle of risk factors; as such, a pure risk allocation approach cannot be applied to alternatives. However, measuring factor exposures of alternative assets can provide valuable information about their risk-return profiles, which should be taken into account when allocations to traditional asset classes are considered. For instance, factor exposure analysis of private equity will highlight the fact that it has significant exposures to size and credit factors. Therefore, the investor may choose to tilt the exposure of the portfolio's traditional assets to other factors.

2.6 CONCLUSION

Chapter 1 introduced the basics of the asset allocation process and studied the mean-variance approach to asset allocation. This chapter extended the concepts discussed in Chapter 1 to take into account practical issues that arise while applying the asset allocation process. This chapter began with a discussion of tactical asset allocation (TAA), with a focus on the costs and benefits of this strategy when applied to portfolios of traditional and alternative investments. It was argued that because the rebalancing of alternative investments is costly, a higher level of skill is needed to make TAA a value-added activity.

Next, the chapter discussed several extensions to the mean-variance approach. The focus here was on those extensions that are important to alternative assets. For instance, we discussed how illiquidity and estimation risk can be taken into account when performing mean-variance optimization. While these extensions may not provide perfect answers, they represent excellent starting points to the asset allocation process and can serve as checks against other, perhaps more heuristic approaches that are adopted by asset allocators.

Risk budgeting and the risk parity approach were discussed next. Risk budgeting is a valuable tool for analyzing a portfolio's risk-return profile and imposing risk constraints desired by asset owners. Risk parity uses the results of risk budgets to create portfolios in which each asset contributes the same amount to the total risk of the portfolio. The chapter discussed the pros and cons of risk parity and pointed out that some of the reasons given in support of risk parity may not apply to alternative investments.

Finally, the chapter discussed factor investing. This is a relatively new topic in the investment community and, similar to other new ideas, has to be evaluated carefully when applied to alternative investments. The primary benefit of factor investing is that it informs investors that returns will come from being exposed to certain risk factors and allows asset owners to decide if earning returns through exposure to these certain risk factors is consistent with their objectives and constraints.

NOTES

1. For the history and theory of TAA, see Lee (2000).
2. See Tokat, Wicas, and Stockton (2007).
3. It is related to the correlation between the recommended weight and the actual weights.
4. See Dahlquist and Harvey (2001); Silva (2006); Tokat, Wicas, and Stockton (2007); Van Vielt and Blitz (2009); Faber (2013); and Hamilton and de Longis (2015).
5. This section is partly based on Tokat, Wicas, and Stockton (2007).
6. The following model is based on Lo, Petrov, and Wierzbicki (2003) and Hayes, Primbs, and Chiquoine (2015).
7. For further discussion, see Fabozzi et al. (2007).
8. For further discussion, see Fabozzi et al. (2007).
9. See Baker, Bradley, and Wurgler (2011) and Asness, Frazzini, and Pedersen (2012).
10. See Frazzini and Pedersen (2014) and Schneider, Wagner, and Zechner (2016).
11. A recent paper by Harvey, Liu, and Zhu (2016) documents the reported discovery of 316 factors. The authors demonstrate that most of these are not reliable sources of risk premium.
12. See Zhang (2005) and Santos and Veronesi (2005).

REFERENCES

Ang, A. 2014. *Asset Management: A Systematic Approach to Factor Investing*. New York: Oxford University Press.

Asness, C., A. Frazzini, and L. H. Pedersen. 2012. "Leverage Aversion and Risk Parity." *Financial Analysts Journal* 68 (1): 47–59.

Asness, C., T. Moskowitz, and L. H. Pedersen. 2010. "Value and Momentum Everywhere." AFA 2010 Atlanta Meetings Paper, March 6. http://ssrn.com/abstract=1363476.

Baker, M., B. Bradley, and J. Wurgler. 2011. "Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly." *Financial Analysts Journal* 67 (1): 1–15.

Dahlquist, M., and C. Harvey. 2001. "Global Tactical Asset Allocation." Duke University Working Paper.

Faber, M. 2013. "A Quantitative Approach to Tactical Asset Allocation." Mabane Faber Research White Paper.

Fabozzi, F., P. Kolm, D. Pachamanova, and S. Focardi. 2007. *Robust Portfolio Optimization and Management*. Hoboken, NJ: John Wiley & Sons.

Frazzini, A., and L. H. Pedersen. 2014. “Betting against Beta.” *Journal of Financial Economics* 111:1–25.

Hamilton, M., and A. de Longis. 2015. “Portfolio Manager Global Multi-Asset Group Dynamic Asset Allocation through the Business Cycle: A Macro Regime Approach.” Oppenheimer Funds White Paper.

Harvey, C., Y. Liu, and H. Zhu. 2016. “… and the Cross-Section of Expected Returns.” *Review of Financial Studies* 29 (1): 5–68.

Hayes, M., J. Primbs, and B. Chiquoine. 2015. “A Penalty Cost Approach to Strategic Asset Allocation with Illiquid Asset Classes.” *Journal of Portfolio Management* 41, no. 2 (Winter): 33–41.

Hibbert, J., A. Kirchner, G. Kretzschmar, R. Li, and A. McNeil. 2009. “Liquidity Premium: Literature Review of Theoretical and Empirical Evidence.” Barrie & Hibbert Research Report version 1.1.

Hull, B., and X. Qiao. 2015. “A Practitioner’s Defense of Return Predictability.” July 22. http://ssrn.com/abstract=2609814.

Idzorek, T., and M. Kowara. 2013. “Factor-Based Asset Allocation vs. Asset-Class-Based Asset Allocation.” *Financial Analysts Journal* 69 (3): 1–11.

Lee, W. 2000. *Theory and Methodology of Tactical Asset Allocation*. New Hope, PA: Frank J. Fabozzi Associates.

Lo, A., C. Petrov, and M. Wierzbicki. 2003. “It’s 11pm—Do You Know Where Your Liquidity Is?” *Journal of Investment Management* 1 (1): 55–93.

Pearson, N. D. 2002. *Risk Budgeting: Portfolio Problem Solving with Value at Risk*. New York: John Wiley & Sons.

Pesaran, H. 2010. “Predictability of Asset Returns and the Efficient Market Hypothesis.” CWPE Working Paper 1033, May.

Santos, T., and P. Veronesi. 2005. “Cash-Flow Risk, Discount Risk, and the Value Premium.” Working paper, Columbia University.

Schneider, P., C. Wagner, and J. Zechner. 2016. “Low Risk Anomalies.” February 1. http://ssrn.com/abstract=2593519.

Silva, H. 2006. “Modern Tactical Asset Allocation.” Proceedings of the CFA Conference on Asset Allocation: Alpha and Beta Investment Strategies, CFA Institute.

Tokat, Y., N. Wicas, and K. Stockton. 2007. “Practical Guidelines for Tactical Asset Allocation Strategy Evaluation.” *Journal of Investing* 16, no. 3 (Fall): 33–41.

Van Vliet, P., and D. Blitz. 2009. “Dynamic Strategic Asset Allocation: Risk and Return Across Economic Regimes.” Robeco Asset Management Working Paper.

Zhang, C., and H. Kazemi. 2015. “Hedge Funds Returns and Market Anomalies.” CISDM Working Paper, University of Massachusetts.

Zhang, L. 2005. “The Value Premium.” *Journal of Finance* 60 (1): 67–103.

CHAPTER 3

The Endowment Model

Investors allocating assets to alternative investments need a framework on which to build their portfolios. What should the size of the allocation to traditional investments be, relative to alternative investments? Within alternative investments, how should the portfolio be diversified across asset classes, styles, and managers? For many, the answers come from a study of the investment practices of the managers of the largest endowment and foundation portfolios.

3.1 DEFINING ENDOWMENTS AND FOUNDATIONS

Endowments refer to the permanent pools of capital owned by institutions such as colleges, universities, hospitals, museums, and religious organizations. When well funded and well managed, an endowment can provide a permanent annual income to the organization, while maintaining the real value of its assets in perpetuity. The idea of perpetuity is not a theoretical concept. The two largest U.S. university endowments are owned by Harvard University and Yale University, which were founded in 1636 and 1701, respectively. Universities that are more than 310 years old with assets of over $23 billion each can operate under the assumption that their assets will exist in perpetuity. The assets held by an endowment can generate income that offsets the impact of economic fluctuations over the course of the business cycle.

Most endowments are run by a single organization but may be funded by thousands of donors. In the United States, each organization is typically organized as a tax-free charity, in which individuals receive a tax deduction for making charitable donations. The investment income of the organization may also be tax exempt. Donations to the organization can be made in many forms, including cash, real estate, or equity securities, as well as art and other collectibles. Noncash donations are frequently sold and the proceeds reinvested according to the strategic asset allocation of the endowment manager. The endowment fund of a single university may be composed of thousands of smaller gifts, many of which are segregated to fund specific scholarships, professorships, or the maintenance of specific buildings or academic programs. These restricted gifts may require that the university maintain the **corpus**, the nominal value of the initial gift, while spending the income generated by the gift to benefit the stated purpose.

Exhibit 3.1 shows the significant asset size of the U.S. and Canadian college and university endowment community. As of June 2014, the National Association of College and University Business Officers (NACUBO) reported that endowment

EXHIBIT 3.1 Assets of the Largest North American University Endowments

	Assets ($ billion) as of June 2008	Assets ($ billion) as of June 2014	Year Founded
Harvard University	$ 36.6	$ 35.8	1636
University of Texas System	$ 16.1	$ 25.5	1883–1895
Yale University	$ 22.9	$ 23.9	1701
Stanford University	$ 17.2	$ 21.4	1891
Princeton University	$ 16.3	$ 21.0	1746
Massachusetts Institute of Technology	$ 10.1	$ 12.4	1861
Total assets of the six largest endowments	$119.2	$140.0	
Total assets of U.S. and Canadian endowments	$412.8	$516.0	
Number of endowments > $1 billion AUM	77	91	

Source: 2015 NACUBO-Commonfund Study of Endowments.

assets totaled $516 billion, with $140 billion held by the six universities with the largest endowment funds. The total value of endowments likely exceeded $516 billion as of June 2014, as this figure includes only those assets of the 832 endowments that responded to the survey. Should nonsurveyed colleges and universities also have endowments, then the total assets would exceed the amount reported in Exhibit 3.1.

The wealth of college and university endowments is highly concentrated at a small number of institutions. As of June 2014, 91 of the 832 colleges and universities reporting to the NACUBO survey had assets exceeding $1 billion. These largest endowments controlled 74% of the total endowment and foundation assets under management (AUM) held by U.S. and Canadian colleges and universities.

Foundations are similar to endowments but tend to differ in a number of ways: (1) foundations are grant-making institutions, whereas endowments tend to be funds established by educational, health-care, or religious organizations; (2) foundations tend to be finite lived, whereas endowments tend to be perpetual; (3) foundations are more subject to minimum spending requirements; and (4) foundations are less likely to be funded from ongoing donations.

Foundations located in the United States have even greater assets than college and university endowments. At the end of 2012, the Foundation Center estimated that U.S. foundations controlled more than $715 billion in assets, the vast majority of which were held by independent, individual, and family foundations (see Exhibit 3.2). The Foundation Center estimates that 22% of the grants made by the top 1,000 foundations in 2012 were awarded to educational charities, 38% to health and human services, 10% to arts and cultural programs, 7% to the environment and animals, and the remaining 23% to charities with various other purposes.

There are a number of different structures for foundations (see Exhibit 3.3). Some are similar to endowments, whereas others differ notably. **Operating foundations** have the greatest similarity to endowments, as the income generated by an endowment is used to fund the operations of the charitable organization. Some of the largest operating foundations are sponsored by global pharmaceutical companies

EXHIBIT 3.2 Assets of the Largest U.S. Foundations

	Assets ($ billion) 2012
Bill & Melinda Gates Foundation	$ 37.1
Ford Foundation	$ 11.2
J. Paul Getty Trust	$ 10.5
Robert Wood Johnson Foundation	$ 9.5
W.K. Kellogg Foundation	$ 8.1
William and Flora Hewlett Foundation	$ 7.7
Total assets of the six largest foundations	$ 84.1
Total assets of U.S. foundations	$715.5
Number of foundations > $2.5 billion AUM	25

Source: The Foundation Center, 2014.

with the goal of distributing medicine to patients who cannot afford to purchase these lifesaving remedies.

Community foundations are based in a specific geographical area, concentrating the charitable giving of the region's residents. The gifts and investment returns received by the community foundation are distributed in the form of grants to other charities in the community. In contrast to endowments and operating foundations, community foundations do not operate their own programs. A single community foundation may partially fund the operations of dozens of charities within a specific region, typically making grants to organizations with a variety of purposes.

Corporate foundations are sponsored by corporations, with gifts provided by the corporation and its employees. Like community foundations, corporate foundations frequently concentrate their financial donations to charities located in the communities where the firm has the greatest number of employees or customers.

Unlike endowments, many foundations find it difficult to survive in perpetuity. In fact, some foundations are designed to last for only a designated period of time. The ability of endowments, operating foundations, and community foundations to solicit gifts greatly increases the probability of the organization's assets lasting into perpetuity.

Most **independent foundations** are funded by an individual or a family. These foundations may be founded by a single gift, often by the senior executive of a large corporation who donates wealth in the form of stock. Donating stock to any charity

EXHIBIT 3.3 Assets, Gifts, and Giving at U.S. Foundations ($ billion)

	2012 Total Assets	2012 Total Gifts Paid	2012 Total Gifts Received
Independent foundations	$584.0	$35.4	$32.2
Corporate foundations	$ 23.1	$ 5.5	$ 4.6
Community foundations	$ 64.9	$ 4.9	$ 7.5
Operating foundations	$ 43.3	$ 6.0	$ 7.8
All foundations	$715.3	$51.8	$52.1

Source: The Foundation Center, 2014.

may provide significant tax benefits. The charitable donation may be tax deductible at the current market value, while capital gains on the appreciated stock position are eliminated for tax purposes by the donation. When tax law allows for this structure, the donors reduce their tax burden in two ways: (1) from the forgiven capital gains taxes on the stock's appreciation, and (2) from the tax deduction on the current market value of the charitable donation.

Independent foundations may present an exceptional challenge for a portfolio manager. First, the wealth of the foundation is often concentrated in a single stock, which increases the idiosyncratic risk of the portfolio. Maintaining this undiversified portfolio can lead to spectacular wealth or gut-wrenching drawdowns, so foundations may wish to reduce the size of the single stock position either rapidly or on a specific schedule. Second, independent foundations do not typically receive gifts from external donors. Once the foundation has been established by the individual or the family donation, many independent foundations do not receive subsequent gifts.

3.2 INTERGENERATIONAL EQUITY, INFLATION, AND SPENDING CHALLENGES

James Tobin stated that the key task in managing an endowment is to preserve equity among generations. The investment goal of an endowment manager should be to maintain **intergenerational equity**, balancing the need for spending on the current generation of beneficiaries with the goal of maintaining a perpetual pool of assets that can fund the operations of the organization to benefit future generations. Stated quantitatively, intergenerational equity may be expressed by a 50% probability of maintaining the real, or inflation-adjusted, value of the endowment in perpetuity. When the probability of the endowment surviving perpetually is low, such as 25%, the current generation has an advantage due to the high spending rate of the endowment. Conversely, a high probability of perpetuity, such as 75%, gives an advantage to future generations, as the endowment would likely survive indefinitely even if the current rate of spending were increased.

The challenge of the endowment manager is to maintain the long-term, inflation-adjusted value of the endowment's corpus, or principal value. The value of the endowment is constantly changing: growing with gifts, falling with spending to fund the organization's mission, and changing with the net returns to the investment portfolio. External forces can also impact the assets and spending of an institution, as gifts, research grants, and governmental funding may change substantially from one year to the next:

$$\text{Change in endowment or foundation value} = \text{Income from gifts} \\ -\text{Spending} + \text{Net investment returns}$$

In fiscal year 2014, NACUBO estimated that the average endowment spent 4.4% of assets, with endowments having assets either above \$1 billion or below \$25 million spending at an average rate of 4.6%. Although endowment spending funded 10.5% of the budget of the average university, 35% of the budgets of Harvard and Yale were recently funded by endowment spending. In contrast to endowments, which typically have flexibility in their **spending rate** (which is the fraction of asset value spent each

EXHIBIT 3.4 Returns of North American University Endowments

	Ending June 2014			
Index	**1 Year**	**3 Years**	**5 Years**	**10 Years**
60% MSCI World Equity Index, 40% Barclays Global Aggregate Bond Index	5.3%	8.6%	10.9%	6.3%
60% Russell 3000 Equity Index, 40% Barclays U.S. Aggregate Bond Index	5.7%	11.3%	13.5%	6.9%
40% Russell 3000 Index, 20% MSCI World Equity Index, 40% Barclays Global Aggregate Bond Index	6.0%	10.0%	12.5%	6.9%
MSCI World Index Free U.S. Currency	6.2%	11.8%	15.0%	7.2%
Russell 3000 Index	6.9%	16.5%	19.3%	8.2%
Barclays Capital Global Aggregate	5.0%	2.7%	4.5%	5.3%
Barclays Capital U.S. Aggregate	3.9%	3.7%	4.9%	4.9%
Endowments over $1 billion	16.5%	9.5%	12.1%	8.2%
Total endowment	15.8%	8.9%	11.8%	7.0%
Consumer Price Index	2.0%	2.4%	1.2%	2.3%
Higher Education Price Index	1.6%	1.9%	1.8%	3.3%

Source: Bloomberg, 2014 NACUBO-Commonfund Study of Endowments.

year), U.S. law requires that foundations spend a minimum of 5% per year on operating expenses and charitable activities. Should charitable contributions received by endowments and foundations decline during times of weak investment returns or rising inflation, the real value of an endowment can fall substantially in a short period of time. Given that foundations have a minimum spending requirement of 5%, while endowments have flexibility in their spending rate, it is easier for endowments to operate in perpetuity than it is for foundations to do so. This is because endowments can reduce their spending rate below 5% of the endowment value during times of crisis.

For an endowment or a foundation to last in perpetuity and provide grants of growing value to its beneficiaries, the returns to its portfolio must exceed the rate of inflation by a wide margin. Exhibit 3.4 shows that the Consumer Price Index (CPI) measure of inflation rose by an annual average of 2.3% in the 10 years ending June 2014. (As of January 2015, the CPI was weighted 32.7% on housing prices, 12.4% on medical and educational expenditures, 22.3% on commodities, and 32.6% on other goods and services.) During the same 10-year period, the Higher Education Price Index (HEPI), a measure of price inflation most relevant to U.S. colleges and universities, rose by an annual average of 3.3%. Salaries of faculty and administrators make up the largest percentage of HEPI at 45.8%; clerical and service employee salaries are weighted at 26.6%; fringe benefits make up 13.1%; and services, supplies, and utilities make up approximately 14.6%. The rate of inflation as measured by HEPI is typically higher than that measured by CPI.

In order to maintain the real value of assets into perpetuity, as well as to meet a payout ratio, a foundation needs to achieve a return target. A **return target** is a level of performance deemed necessary to satisfy the goals of the owners or beneficiaries of the associated assets. A foundation that has a spending rate of 5%, does not have any regular gift income, and wishes to preserve the real value of its assets has an aggressive return target: the rate of inflation plus 5%, or even higher when the foundation's spending rate exceeds 5%. When measured relative to CPI inflation over the prior 10 years, this return target is 7.3%, a rate higher than the returns to a blended stock and bond portfolio over the same period. Return targets are even higher when the institution faces a higher inflation rate, such as HEPI, but lower when there is a substantial and regular flow of donations to the organization. For institutions without substantial gifts, endowment values likely declined in real terms over the previous decade due to investment returns that fell short of the targeted return of inflation plus 5%.

David Swensen, the chief investment officer of Yale University and author of *Pioneering Portfolio Management* (2009), challenges endowment managers to resist the temptation to increase spending rates after periods of extremely high returns. He argues that limiting spending will allow an endowment to better survive cyclical drawdowns and better compound wealth in perpetuity.

3.3 THE ENDOWMENT MODEL

Aggressive return targets, as well as the perpetual life of many endowments and foundations, have led to an equally aggressive asset allocation. The asset allocation of major endowments and foundations, which typically includes substantial allocations to alternative investments, has been called the **endowment model**.

Universities with large endowments have been early adopters of alternative investments, well known as sophisticated investors across all areas of alternative investments. The financial success of these investors has been much discussed, even spawning a subset of investors who seek to earn higher returns by following similar investment strategies. Many credit the endowment model, or at least the most articulate description of the endowment model, to David Swensen. Most of the U.S. colleges and universities with endowment assets in excess of $1 billion tend to invest large portions of their endowment portfolios in alternative investments, following the example of Yale University.

The six largest endowment funds, as shown in Exhibit 3.1, suffered substantial drawdowns and liquidity issues in 2008 and 2009, as assets declined from $119.2 billion in June 2008 to $95.0 billion in June 2010. After a $31.8 billion (–26.7%) loss in value in the year ending June 2009, including spending and gifts, the endowment model attracted some criticism. However, Exhibit 3.4 shows that in the 10 years ending June 2014, endowments with assets exceeding $1 billion earned returns of 8.2%, far surpassing the returns to domestic equity markets or a 60% domestic equity/40% domestic fixed-income allocation, which is the traditional benchmark for institutional investors. Ten-year annualized returns of 8.2% refute the idea of a lost decade, rendering the three-year drawdown of an annualized –3.5% from 2008 to 2010 relatively benign, at least when viewed in hindsight.

According to the 2014 NACUBO-Commonfund Study of Endowments, the world's largest college endowments continue to increase their allocations to alternative investments. Year after year, allocations to domestic equities decline while allocations to alternative investments and international equities increase. As of June 30, 2014, the latest time period available, these endowments held 13% in U.S. equities,

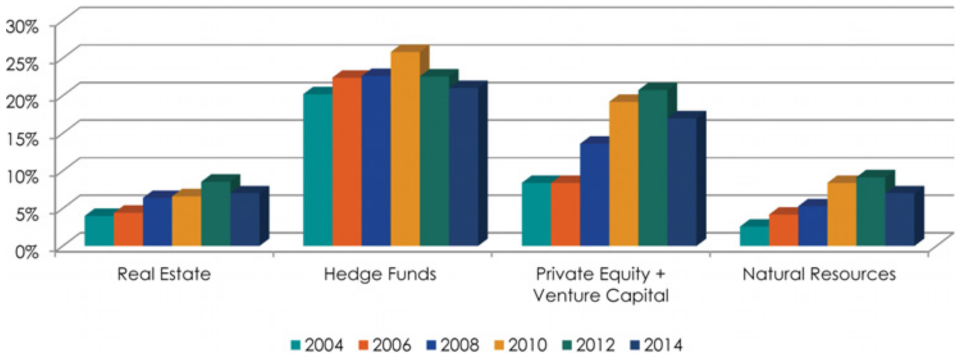


EXHIBIT 3.5 Exposure to Alternative Investments within Large Endowment Funds
Source: NACUBO, 2014.

18% in international equities, 12% in fixed income and cash, and 57% in alternative investments. As shown in Exhibit 3.5, alternative asset allocations have exploded since 2004: hedge fund allocations grew to 21%; private equity and venture capital allocations increased to 17%; and natural resources and real estate experienced dramatic increases, to 14% of endowment portfolio holdings. These highly successful investors now allocate more to international equities than to U.S. equities, more to hedge funds than to fixed income, and more to private equity and venture capital than to domestic public equities. The largest endowments have increased their allocations to alternative investments from 32.5% in 2002 to 57% in 2014.

Swensen believes strongly in an equity orientation, seeking to participate in the ownership of both public and private equity securities and real assets. The role of fixed income is to provide liquidity and a tail hedge that serves to reduce potential losses in the portfolio. Yale University chooses not to invest in either investment-grade or high-yield bonds due to the inherent principal–agent conflict. As Swensen explains the conflict, corporate management explicitly works for stockholders and may choose to make decisions that benefit stockholders, even when those decisions are to the detriment of bondholders. Given this conflict and the fact that the total returns to corporate bonds are less than 1% above government bonds on a long-term, net-of-defaults basis, the incremental return to corporate bonds may not warrant inclusion in the endowment portfolio. While sovereign bonds provide liquidity and a tail hedge in a time of crisis, corporate bonds can experience a reduction in liquidity and a disastrous loss of value during extreme market events, producing the opposite effect to what fixed income should have on a portfolio. Similarly, foreign bonds are not held in the Yale University portfolio because, while the return may be similar to that of domestic sovereign bonds, the addition of currency risk and unknown performance during times of financial crisis is not consistent with Swensen's goals for the fixed-income portfolio.

Not all endowments have a similar affinity toward alternative investments. Exhibit 3.6 shows the asset allocation of the equally weighted endowment, which averages asset allocation across all 832 endowments surveyed by NACUBO, ranging

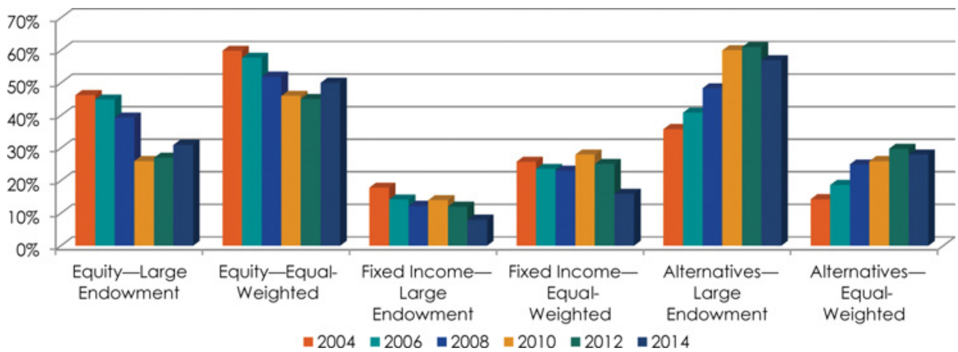


EXHIBIT 3.6 Asset Allocation of Large versus Average College and University Endowments
Source: NACUBO, 2014.

from those with assets below $25 million to those over $1 billion. While the average endowment has a smaller allocation to alternatives than the largest endowments, the equal-weighted average allocation to alternative investments more than doubled (from 11.8% to 28%) between 2002 and 2014. In fact, the allocation to alternative investments at college and university endowments increases monotonically with asset size: endowments between $25 million and $50 million have a larger allocation than those below $25 million, while those with $100 million to $500 million in assets have larger allocations than endowments with assets between $50 million and $100 million. Returns over the past 10 years reflect the same patterns: the largest endowments have both the highest returns as well as the largest allocations to alternative assets.

3.4 WHY MIGHT LARGE ENDOWMENTS OUTPERFORM?

Investors worldwide, from pensions, endowments, and foundations to individual investors, have become attracted to the endowment model, seeking to emulate the returns earned by the largest endowments over the past 20 years. However, evidence shows that not all aggressive allocations toward alternative investments necessarily earn similar returns. This may be due to key advantages particular to large endowments. The literature discusses at least six such advantages that may explain the excellent returns earned by large endowments in recent years:

1. An aggressive asset allocation
2. Effective investment manager research
3. First-mover advantage
4. Access to a network of talented alumni
5. Acceptance of liquidity risk
6. Sophisticated investment staff and board oversight

Investors lacking these advantages may find it difficult to earn top returns, even when following the endowment model.

3.4.1 An Aggressive Asset Allocation

In the world of traditional investments, a number of studies—including those by Ibbotson and Kaplan (2000); Brinson, Hood, and Beebower (1986); and Brinson, Singer, and Beebower (1991)—show that the strategic asset allocation of pension plans accounts for between 91.5% and 93.6% of the variance in fund returns. The remaining portion of the variance in fund returns, just 6.4% to 8.5%, can be explained by security selection and market timing. (Note that it was the variance in returns that was measured rather than the amount of returns.)

$$\begin{aligned}\text{Return attribution} &= \text{Contributions from strategic asset allocation}\\ &\quad + \text{Security selection} + \text{Market timing/tactical allocation}\end{aligned}$$

The returns from strategic asset allocation are measured by multiplying the targeted long-term asset allocation weights by the benchmark returns to each asset class. Security selection is defined as the return within asset classes relative to a benchmark, such as the return to the domestic fixed-income portfolio when compared to the domestic fixed-income benchmark. **Market timing** is measured as the return earned from the variation of asset class weights versus the policy or target asset class weights. Value is added from market timing, or tactical asset allocation, when the returns to overweighted asset classes outperform the returns to underweighted asset classes. For example, when the actual equity allocation is 42% and the target equity allocation is 40%, the return to market timing is positive when the equity index outperforms the returns of the other asset classes in the portfolio.

Swensen (2009) explains the role of tactical asset allocation and rebalancing. Investors are encouraged to be contrarian and consider valuation when making allocations. Rather than considering what his peers were doing, Swensen entered alternative asset classes earlier and more aggressively than did other institutional investors. In contrast to the common practice of increasing allocations to asset classes after a period of outperformance, Swensen sought to aggressively **rebalance** (i.e., transact so as to cause portfolio weights to return to prespecified values) to strategic asset allocation weights by selling outperforming asset classes and buying underperforming ones. This can be psychologically difficult, as it requires buying equities during a crash and selling certain assets when other investors are clamoring to increase their allocations to those assets. Market timing between risky and less risky asset classes, such as equities versus cash, can be dangerous due to the large difference in expected returns. While generally eschewing market timing, Swensen suggests tilting the portfolio toward undervalued assets and away from overvalued assets between asset classes with similar long-term return expectations, such as real estate and natural resources. Rebalancing can add significant value, so Swensen sought to aggressively rebalance in real time. In a volatile year, such as 2003, Yale's rebalancing activity added as much as 1.6% to annual returns.

Brown, Garlappi, and Tiu (2010) analyze the returns to endowment funds over the period 1984 to 2005 and find a much different story than that seen in prior literature regarding return attribution in pension plans. This study shows that just 74.2% of the returns can be explained by the endowment's strategic asset allocation. With market timing explaining 14.6% of returns, and security selection explaining

8.4%, endowment funds show a much larger contribution to returns from dynamic asset allocation and manager selection.

While it can be easy to replicate the asset allocation of endowment funds, investors seeking to emulate the success of the endowment model will find it much more difficult to profit from market timing and security selection. One reason may be that while pension plans are seen to focus largely on passive allocations within each asset class, endowment funds place a much greater emphasis on manager selection. Brown, Garlappi, and Tiu also find that the return from policy allocation is quite similar across endowments. The contribution from asset allocation explains just 15.3% of the return differences across endowments, while selection explains 72.8% and timing contributes just 2.5%. The endowment model, then, does not seem to be an asset allocation story but rather a story of superior manager selection.

3.4.2 Effective Investment Manager Research

From a risk-budgeting perspective, many endowment managers prefer to spend the majority of their active risk budget in alternatives. On average, these large investors allocate to eight managers within traditional asset classes and more than 11 managers in alternative asset classes. Endowments have exploited both their networks of successful alumni and their first-mover advantage in allocating to the best-performing managers, many of whom have now closed their funds to new investors. While common wisdom assumes that the sole secret to endowment success is the large allocation to alternative investments, the top endowments enhance performance further by allocating to managers who outperform. In fact, the largest endowments have historically outperformed in nearly every asset class, both traditional and alternative. Those investors seeking to replicate the success of endowments should be cautioned that although this outperformance within asset classes can add up to 2% per year to performance, it is unlikely to be replicated by an alternatives-heavy allocation if investors lack the talented staff and valuable network of invested managers that many endowments have cultivated.

An examination of university annual reports shows that in addition to an alternatives-heavy asset allocation, which can enhance returns and reduce risk, many universities have superior security selection and manager selection skills. This skill, or high risk tolerance, is seen when examining the return in each asset class portfolio compared to that asset class's benchmark. Harvard's 2014 annual report shows that returns beat asset class benchmarks annually over five years, including 1.9% in real assets and 3.9% in absolute return, while falling 0.5% behind the private equity benchmark. Even in the asset classes considered to be most efficiently priced, such as publicly traded equity and fixed income, Harvard outperformed annually by 0.6% and 2.9%, respectively, over the same period. In total, Harvard's entire endowment portfolio beat the return on the asset-weighted benchmark by 1.4% annually, adding more than $1 billion in excess returns in just five years.

Swensen (2009) demonstrates the importance of manager selection within the alternative investment universe. In liquid, efficient markets, the dispersion of returns across asset managers is relatively small. For example, in U.S. fixed income over the 10-year period ending in 2005, there was a mere 0.5% difference in returns between managers at the first and third quartiles of returns. Equity markets have return dispersion across managers of between 1.9% and 4.8%. Pension plans often invest in passive index funds rather than with active managers in traditional asset classes. However, index funds are generally unavailable in most alternative investment asset classes. Siegel (2008) cites literature that documents the difference in annual returns between managers ranked at the 25th and 75th percentiles. These numbers are relatively small in traditional assets, with 0.5% in fixed income, 2.7% in U.S. equity, and 3.9% in non-U.S. equity.

In contrast, the value added by active managers in alternative investments can be quite substantial. In inefficient markets, managers have a greater opportunity to profit from skill, information, and access to deal flow. Dispersion in alternative investments is much higher, with 7.4% in hedge funds, 14.2% in private equity buyouts, and 35.6% in venture capital. In many cases, especially in private equity, investments are not attractive when investing in the median manager. In order for a private equity investment to outperform public equity on a risk-adjusted basis and adequately compensate for the liquidity risk in these investments, investors need to allocate to managers who deliver returns far above the median manager in each asset class.

3.4.3 First-Mover Advantage

It appears that the largest endowments have significant skill in selecting the top-performing managers within each asset class. Lerner, Schoar, and Wang (2008) explain that this ability to select top managers may be related to the **first-mover advantage** (i.e., benefits emanating from being an initial participant in a competitive environment): large endowments invested in many alternative asset classes years earlier than pension funds and smaller endowments did, and may therefore have an advantage. For example, Takahashi and Alexander (2002) explain that Yale University made its first investments in natural resources in 1950, leveraged buyouts in 1973, venture capital in 1976, and real estate in 1978. In contrast, Lerner, Schoar, and Wang explain that corporate pensions began investing in venture capital only in the 1980s, while public pension plans did not make their first venture capital investments until the 1990s.

Many of the funds of managers who have earned top-quartile performance in these asset classes have been closed to new investors for many years. Newer investors seeking access to top managers in alternative investment asset classes, especially in venture capital, are destined to underperform when the top managers allow commitments only from those investors who participated in their earlier funds.

Lerner, Schoar, and Wongsunwai (2007) show that endowments earn higher average returns in private equity, likely due to the greater sophistication of their fund-selection process. Endowment funds have higher returns than do other investors when making allocations to first-time private equity fund managers. Once an endowment fund has become a limited partner in a private equity fund, it seems to be more efficient at processing the information provided by each general partner. The follow-on funds that endowments select for future investment outperform funds to which endowments decline to make future commitments.

Mladina and Coyle (2010) identify Yale University's investments in private equity as the driving factor in the endowment's exceptional performance. It can be difficult to emulate Yale's outperformance in private equity and venture capital investments, as its venture capital portfolio has earned average annual returns of 31.4% since its inception through fiscal year 2007. In fact, this study suggests that without the private equity and venture capital investments, the returns to the Yale endowment would be close to that of the proxy portfolio.

3.4.4 Access to a Network of Talented Alumni

Perhaps the first-mover advantage and the manager-selection skill of top endowments can be attributed to the superior network effect. An institution has a positive **network effect** when it has built relationships with successful people and businesses that may be difficult for others to emulate. Alumni of the universities in Exhibit 3.1 are noted for being among the most successful U.S. college graduates, in terms of both academics and business. As measured by scores on the SAT® exam,[1] Harvard, Yale, Stanford, Princeton, and the Massachusetts Institute of Technology routinely select from the top 1% to 5% of students. In 2003, the median SAT score for all college-bound students, including both verbal and mathematics scores, was approximately 1,000. Top universities attract students with average scores exceeding 1,400. Graduates of these schools also tend to have the highest initial and midcareer salaries.

A study by Li, Zhang, and Zhao (2011) correlated manager-specific characteristics to the returns of the hedge funds they managed. In contrast to the median SAT score of all college-bound students of 1,000, Li, Zhang, and Zhao found that the middle 50% of hedge fund managers attended colleges and universities with average SAT scores between 1,199 and 1,421 (the 79th and 97th percentiles, respectively), demonstrating that the majority of hedge fund managers attended the most competitive colleges and universities. Within the group of studied hedge fund managers, the research showed that those who attended undergraduate colleges with higher average SAT scores have higher returns and lower risk. For example, a 200-point difference in SAT scores, such as that between 1,280 and 1,480, was correlated with higher annual returns of 0.73%. Not only did managers who attended top universities have higher returns, but they did so at lower risk and earned greater inflows during their tenures as fund managers. The authors suggest that talented managers are attracted to hedge funds due to the incentive fee structure, which rewards performance over asset gathering. In contrast to their studies on hedge fund managers, Li, Zhang, and Zhao found that SAT scores did not seem to affect the asset gathering or excess returns earned by mutual fund managers, as these managers are compensated for gathering assets, not for earning excess returns.

Many alumni of top universities wish to continue an association with their alma mater, the university from which they received their undergraduate degree. The ability of top endowment funds to outperform can be perpetuated by this important network of relationships, to the extent that these talented professionals either choose to work for the university's endowment fund or guarantee the endowment investment access to the funds they manage.

Barber and Wang (2013) show that the strong returns earned by endowments are directly traced to the size of their alternative investment exposure. Alphas earned by Ivy League schools exceed 3%, while 30 other schools with top SAT scores earned an alpha exceeding 1.7%. There is a reliably positive alpha and return spread between schools with top SAT scores and schools with average scores.

3.4.5 Acceptance of Liquidity Risk

Endowments have a perpetual holding period. With low spending rates and limited liabilities, endowments have a much greater tolerance for risk, including liquidity risk. When viewed in light of the age of leading universities, which for Harvard and Yale now surpasses 300 years, the 10-year lockup period of private equity vehicles appears relatively short term. As the longest-term investors, charged with protecting the real value of endowment principal for future generations of students, universities are seeking to earn **liquidity premiums**, which are higher returns earned by investing in less liquid assets that require long lockup periods. The idea is that the perpetual nature of endowments allows them to easily handle this liquidity risk. Anson (2010) estimates the liquidity premium for private equity at 2% and for direct real estate at 2.7%, while other studies estimate liquidity premiums as high as 10%.

Swensen (2009) explains that less liquid investments tend to have greater degrees of inefficient pricing. On average, investors overvalue liquid assets, which leaves undervalued and less liquid assets for investors with long-term investment horizons. Investors making commitments to long-term assets, such as private equity and private real estate, know that these investments are typically held for 10 years or longer and so require a significant due diligence process before making such a long-term commitment. Investors in more liquid asset classes may not take their investments as seriously, knowing that the investment may be exited after a short-term holding period. Investments that appear to be liquid in normal markets may have constrained liquidity during times of crisis, which is when liquidity is most valued.

3.4.6 Sophisticated Investment Staff and Board Oversight

All investors need a process by which asset allocations are set and managers are selected. Traditionally, an institutional investor would have an internal staff that would make recommendations to an investment committee, which would then vote on recommendations at quarterly meetings. The quality of the votes and recommendations depends on the experience and composition of the members of the endowment's staff and investment committee.

Investors with smaller assets under management tend to have smaller staffs. In 2011, NACUBO estimated that college and university endowments with less than \$100 million had just 0.4 staff members dedicated to endowment issues, meaning that a single staff member, such as the chief financial officer or treasurer, was responsible for the endowment along with a wide variety of other budget and financial issues. In contrast, the endowments with over \$1 billion in assets tend to have large and sophisticated internal teams, averaging over 10 investment professionals. These teams tend to be well experienced and highly compensated, allowing them to manage some of the assets in-house as well as recommend investment managers. Whereas 79% of the largest endowments employ a chief investment officer (CIO), less than 3% of endowments with less than \$100 million in assets employ someone whose full-time role is to oversee the endowment portfolio.

In addition to internal staff and an investment committee, many endowments employ external consultants. In 2011, NACUBO estimated that 79% to 94% of endowments with between \$25 million and \$1 billion in assets, and 68% of the largest endowments, employed consultants. A **non-discretionary investment consultant** makes recommendations to the endowment on asset allocation, manager selection, and a wide variety of other issues, but leaves the ultimate decision to a vote of the investment committee. There is growing use of the **outsourced CIO** (OCIO) model, in which the endowment gives discretionary authority to an external consultant who may make and implement prespecified decisions, such as manager selection and asset allocation decisions, without taking those decisions to a vote. Endowments with smaller internal teams appear to find the outsourced CIO model attractive, as between 42% and 62% of endowments with assets below $100 million had hired OCIOs by 2011. The trend toward hiring OCIOs accelerated after the 2008 financial crisis, when investors realized that tighter risk controls and quicker rebalancing decisions were needed. Williamson (2013) reports that global OCIO assets under management had reached $1.066 trillion by 2013, including $619 billion in the United States, a growth rate of 59% in just two years. In addition to small endowments and foundations, corporate pensions with liability-driven investing targets are increasingly likely to hire an OCIO.

The Commonfund Institute (2013) notes a number of benefits to hiring an OCIO, especially for endowments that are devoting ever-larger allocations to alternative investments. An OCIO firm will have a large staff and significant infrastructure resources that are shared across all its clients. This institutional-quality firm has resources that could not be afforded by smaller investors. There are economies of scale in manager research, as hedge fund and other alternative managers can visit the consultant or OCIO firm rather than visiting the dozens of underlying investors. The OCIO firm can be cost-effective when compared to attracting, training, and retaining investment professionals, who may be difficult to find and retain in a market where there is a growing demand for those who have experience managing foundation and endowment assets. For investors who do have staff, the OCIO may help train and educate that staff. Whereas 44% of investment decisions take more than three months when an investment committee retains discretion, the OCIO model can make investment and rebalancing decisions on a far more frequent basis.

Lord (2014) studied the common factors shared by the largest and most successful endowments. Ideally, the investment committee would be staffed by investment professionals and others who have experience serving as corporate executives or board members. If those investors have experience in alternative investments and a wider variety of investment strategies, the resulting portfolio tends to be more diversified and experience higher risk-adjusted returns. Investment committees with significant representation from donors or employees of the universities tend to have lower allocations to alternative investments. Decision-making is improved when committee members have multiple perspectives and an ability and willingness to openly debate issues. When adding new members to the investment committee, endowments should seek members with knowledge and experience that differ from those of current committee members. Finally, top-performing endowments have a commitment to educating staff and committee members on new asset classes before allocations are made.

3.5 RISKS OF THE ENDOWMENT MODEL

When applied by the largest investors, the endowment model has created impressive returns over the past 20 years. However, this style of portfolio management comes with a special set of risks. First, portfolio managers need to be concerned about the interactions among spending rates, inflation, and the long-term asset value of the endowment. Second, a portfolio with as much as 60% invested in alternative assets raises concerns of liquidity risk and the ability to rebalance the portfolio when necessary. Finally, portfolios with high allocations to assets with equity-like characteristics and low allocations to fixed income require the portfolio manager to consider how to protect the portfolio from tail risk, which is a large drawdown in portfolio value during times of increased systemic risk.

3.5.1 Spending Rates and Inflation

There is an important tension between the spending rate of the endowment, the risk of the endowment portfolio, and the goal of allowing the endowment to serve as a permanent source of capital for the university. When the endowment fund generates high returns with a low spending rate, the size of the endowment fund increases. This may lead to concerns about intergenerational equity, as the spending on current beneficiaries could likely be increased without compromising the probability of the endowment continuing into perpetuity. Conversely, a conservative asset allocation with a high spending rate may favor the current generation yet imperil the real value of the endowment in the long run.

Kochard and Rittereiser (2008) present a history of endowment spending models. From the founding of the Harvard University endowment in 1649 until the 1950s, endowments were typically managed conservatively, with a focus on earning income from a fixed income–dominated portfolio. It was clear what the spending rate should be, as the entire portfolio yield was typically paid out to support the programs of the university. When bonds matured at face value, the notional value of the corpus was maintained. This sounds relatively straightforward, as the current beneficiaries received a strong income, and the requirement of maintaining the nominal value of gifts made to the foundation was easily achieved. Unfortunately, this conservative asset allocation earned little in the way of real returns, and the focus on minimizing drawdowns led to a low total return and stagnation of the nominal value of endowments.

Between the 1950s and 1970s, this conservatism began to fade. Rather than spending the income generated on the portfolio, endowment managers came to embrace the concept of total return. Consider a fixed-income portfolio with a 5% current yield, which allowed a 5% spending rate. Though the yield was high, the nominal return to the portfolio, net of spending, was zero. In this scenario, the spending needs were easily met, yet the value of the endowment declined in real terms, as the goal of maintaining the real, or inflation-adjusted, value of the corpus was not achieved. In seeking to maximize the portfolio's risk-adjusted return, a **total return investor** (i.e., an investor who considers both income and capital appreciation as components of return) may realize that a 5% current yield is not needed in order for the endowment to have a spending rate of 5%. Moving from a portfolio dominated by fixed income to one with a healthy mix of equity investments may reduce the yield to 3% while increasing total return to 7.5%. With a total return of 7.5%, including income and capital appreciation, the endowment can afford a spending rate of 5% while maintaining the real value of the portfolio, as the 2.5% return in excess of the spending rate can be used to offset the impact of inflation. In order to generate 5% spending, the entire income of 3% is spent, and 2% of the portfolio is sold each year to meet the spending rate.

The investing behavior of endowment managers began to change between 1969 and 1972. The 1969 Ford Foundation publication *Managing Educational Endowments* suggested that endowment portfolios had previously invested with overly conservative asset allocations. Endowment managers were chastised for building underperforming portfolios that were underweighted in equities, not celebrated for serving the university by reducing the risk of loss. The publication proposed that trustees might have imperiled the university by forgoing higher returns that could have increased both the income and the corpus of the endowment. The 1972 Uniform Management of Institutional Funds Act subsequently allowed endowment managers to consider total return when setting the spending rate. The act also allowed the use of external investment managers and encouraged trustees to balance the long-term and short-term needs of the university.

It was during the bull market in public and private equities, from 1982 to 2000, that endowment managers dramatically increased their equity exposures. This produced extremely strong returns that boosted the values of endowments by large multiples, net of spending. The post-2000 drawdown in equity markets led endowments to have an even stronger focus on alternative investments, moving to increase exposure to hedge funds and other assets with lower correlations to equities while reducing exposure to publicly traded equities.

Once an asset allocation is determined, a spending rule must also be established. The earliest endowments spent income only, which tilted the portfolio toward income-producing securities. Later, endowments moved toward spending at a fixed percentage of the current value of the endowment, such as 4%. This fixed spending rate, however, created volatility in the amount of income available to the university. In a year when the return and gifts received by the endowment generated a 20% increase in endowment value, the income to be spent also increased by 20%. Conversely, during a 20% drawdown, the spending rate of the university was slashed by a significant amount. A sticky spending rate, such as $3 million per year, provides certainty of income to the university but can create concerns of intergenerational equity after a large gain or loss in the value of the endowment.

Recognizing that volatility in the income provided to university operations was unwelcome, more flexible spending rules were developed. Spending 4% of the average value of the endowment over the trailing three to five years creates a smoothing process that dampens the impact of the volatility of portfolio returns on the income provided to the university.

David Swensen developed a spending rule for Yale University: each year, the endowment could spend at a rate equal to 80% of the prior year's dollar spending plus 20% of the endowment's long-term spending rate (4.5%). This formula incorporates the prior 10 years of endowment value into the spending rate calculation, providing a stronger smoothing effect than a simple moving average rule.

Swensen (2009) expresses concern about the impact of inflation, market volatility, and high spending rates. Some 70% of endowments use spending rates of between 4% and 6%, with 5% being a popular choice. Yale has set its long-term spending rate at 4.5%, as simulations show that, based on a five-year average of endowment values, a 5% spending rule has a 50% probability of losing half of the endowment's real value at some point over the course of a generation. Using a longer averaging period and a lower spending rate reduces the probability of this disastrous decline in the real value of the endowment portfolio.

Inflation has a particularly strong impact on the long-term real value of university endowments. Ideally, endowments should seek to maintain the real value of the corpus rather than the legal requirement of the notional value. When maintaining the real value of the corpus, long-term spending rates can keep up with inflation; when maintaining the nominal value of the corpus, long-term spending has an ever-declining value in real terms.

The focus on inflation risk has led many endowments to increase allocations to real assets in recent years. Between 2002 and 2014, the largest endowments increased their average allocations to real estate and natural resources from 6% to 14%. Interestingly, while endowments invested a greater portion in real estate than in natural resources in 2002 (4.3% versus 1.7%), they are currently investing the same amount in both natural resources and real estate (7.0%). Real asset investments include inflation-linked bonds; public and private real estate investments; commodity futures programs; and both direct and private equity fund investments in mining, oil and gas, timber, farmland, and infrastructure. Ideally, the real assets portfolio would earn long-term returns similar to those of equity markets, with yields similar to those of fixed income, while experiencing low volatility and low correlation to the fixed income and publicly traded equity assets in the portfolio, as well as higher returns during times of rising inflation.

A report from Alliance Bernstein (2010) calculated the inflation betas of several asset classes. An **inflation beta** is analogous to a market beta except that an index of price changes is used in place of the market index, creating a measure of the sensitivity of an asset's returns to changes in inflation. Just a few assets demonstrated a positive inflation beta, where the assets act as an effective inflation hedge. The majority of assets have a risk to rising inflation; that is, they have a negative inflation beta. According to Alliance Bernstein, commodity futures offered the greatest inflation beta at 6.5, whereas farmland had a beta of 1.7. In the fixed-income sector, 10-year Treasury Inflation-Protected Securities (TIPS) had a beta of 0.8, whereas three-month Treasury bills had a beta of 0.3. Equities and long-term nominal bonds had a strong negative reaction to inflation, with the Standard & Poor's (S&P) 500 Index exhibiting an inflation beta of –2.4 and 20-year U.S. Treasury bonds suffering returns at –3.1 times the rate of inflation. Within equities, small-capitalization stocks had an even greater risk to rising inflation. Companies with lower capital expenditures and fewer physical assets also had a stronger negative response to rising inflation.

3.5.2 Liquidity Issues

In the aftermath of the 2008 financial crisis, many pension funds and endowments have begun to reevaluate their asset allocation policies and, in the process, are paying increased attention to their risk and liquidity management practices. Liquidity represents the ability of an entity to fund future investment opportunities and to meet obligations as they come due without incurring unacceptable losses. These obligations include the annual spending rate as well as the capital calls from private equity and real estate limited partnerships. If there are mismatches between the maturity of an entity's assets and its liabilities, the entity is exposed to liquidity risk.

While liquidity is certainly a risk for endowments, these funds have long lives and can afford to take a fair amount of liquidity risk. Studies (Aragon 2004; Khandani and Lo 2011; Sadkay 2009) have shown that the illiquidity premium is generally positive and significant, ranging from 2.74% to 9.91% for some investment strategies. Ang and Kjaer (2011) suggest that investors should demand steep premiums to bear liquidity risk, as holding less liquid assets may cause investors to forfeit the lucrative opportunity to buy assets at distressed prices during a crisis. For 10-year lockups, this premium may be 6%, while two-year lockups require a 2% premium. The size of this premium varies through time, with studies suggesting that illiquidity premiums declined in the years leading to the financial crisis. Therefore, similar to management of other risks, a portfolio manager has to consider carefully the trade-off between the liquidity risk and the illiquidity premium in determining the size of the illiquid assets in the overall portfolio. These studies also show that, everything else being the same, funds with long lockup periods generally provide a higher rate of return to investors. A long lockup period is a vital tool employed by managers to reduce the cost of liquidity risk. During the most recent financial crisis, funds with long lockup periods were not under pressure to sell their assets at distressed prices. It is important to note that if the underlying assets of a fund are less liquid than the liquidity provisions it offers to its investors, the cost of liquidity risk will increase for all investors, even if only a small fraction of the fund's investors decide to redeem their shares during periods of financial distress. The fact that some pension funds and endowments have decided to reduce their allocations to illiquid assets may signal that the illiquidity premium will be higher in the future. Pension funds and endowments cannot afford to ignore such an important source of return if they are to meet the needs of their beneficiaries.

Effective liquidity risk management helps ensure the ability of a pension fund or an endowment to meet its cash flow obligations, which may not be completely predictable, as they are affected by external market conditions. Due to lack of effective liquidity risk management, many funds experienced severe liquidity squeezes during the latest financial crisis. This forced some to sell a portion of their illiquid assets at deep discounts in secondary markets, to delay the funding of important projects, and, in certain cases, to borrow funds in the debt market during a period of extreme market stress. These experiences have led some to question the validity of the so-called Yale model of pension and endowment management and, in particular, to discourage pension funds and endowments from allocating a meaningful portion of their portfolios to alternative assets.

In *The Global Economic System*, Chacko et al. (2011) explain that liquidity risk rises during a crisis, as declining liquidity and rising volatility increase bid-ask spreads and reduce trading volumes. The book notes that alternative investments have a very high liquidity risk, with private equity, venture capital, real estate, hedge funds, and infrastructure exhibiting liquidity betas in excess of 1.0. While these investments tend to have higher returns over long periods of time, the underperformance during times of crisis can be substantial due to the large exposure to liquidity risk. Chacko et al. also discuss **liquidity-driven investing**, an investment approach emphasizing the role of the liquidity of investments and the time horizon of the investor in the asset allocation decisions. Tier 1 assets are invested in short-term fixed income; tier 2 assets are invested in risky, liquid assets, such as stocks; and tier 3 assets are both risky and illiquid, such as investments in private equity and hedge funds. The endowment should estimate the spending and capital calls for the next 10 years and invest those assets exclusively in tier 1 and tier 2 assets, which can be liquidated quickly at relatively low cost. Tier 3 assets are designed as long-term investments; as such, the size of this allocation should be designed to prevent the need to liquidate these assets in the secondary market before maturity.

One measure of liquidity risk is the sum of the endowment's allocation to private equity and real estate partnerships combined with the potential capital calls from commitments to funds of more recent vintage. Bary (2009) reports, "At Harvard, investment commitments totaled $11 billion on June 30, 2008; at Yale, $8.7 billion, and Princeton, $6.1 billion. These commitments are especially large relative to shrunken endowments. Harvard's endowment could end this month in the $25 billion range; Yale's is about $17 billion, and Princeton's, $11 billion, after investment declines, yearly contributions to university budgets and new gifts from alumni and others."

Takahashi and Alexander (2002) from the Yale University endowment office discuss the importance of understanding the capital call and distribution schedule of private equity and real estate investments. In these private investment vehicles, investors commit capital to a new fund, and that capital is contributed to the fund on an unknown schedule. A typical private equity or real estate fund will call committed capital over a three-year period, focus on investments for the next few years, and then distribute the proceeds from exited investments in years 7 to 12 of the partnership's life. Once an alternative investment program has matured, it may be possible for distributions from prior investments to fully fund capital commitments from new partnerships.

However, when starting a new program, it can be challenging to accurately target the allocation of contributed capital to these long-term partnerships. One rule of thumb is to commit to 50% of the long-term exposure, such as a $10 million commitment once every three years to reach a long-term allocation of $20 million. Takahashi and Alexander offer specific estimates for the speed at which committed capital is drawn down for a variety of different fund types. Real estate funds may draw down uncalled capital at the fastest rate, with an estimate of 40% of uncalled capital to be drawn each year. Venture capital is slower, with 25% the first year, 33.3% the second year, and 50% of the remaining capital called in each subsequent year. Leveraged buyout funds may require a 25% contribution in the first year, with 50% of the remaining capital called in each subsequent year. Notice that not all committed capital is eventually called, so some investors may implement an **overcommitment strategy** by making capital commitments in excess of the targeted investment amount.

During the 2008 crisis, it became very difficult for managers to exit investments, as private equity funds could not float initial public offerings, and real estate funds could not sell properties. As a result, distributions were much slower than expected. When distributions slowed and capital calls continued, some endowments and foundations found it challenging to meet their commitments, as they had previously assumed that the pace of distributions would be sufficient to fund future capital calls. The price of a missed capital call can be steep (up to as much as a forfeiture of the prior contributed capital) and can result in being banned from participating in future funds offered by the general partner.

One feature of the endowment model is the minimal holdings of fixed income and cash. For example, going into 2008, Yale's target for fixed income was 4%, with leverage creating an effective cash position of –4%. Princeton had a combined weight of 4%, and Harvard held approximately 8%. Although income from dividends, bond interest, and distributions from private funds added to the available cash, in many cases the income, fixed income, and cash holdings were not sufficient to meet the current year's need for cash. With a 5% spending rate, it became necessary for these endowment funds to either borrow cash or sell assets at fire-sale prices in order to guarantee the university sufficient income to fund its operations. To the extent that the endowment also had capital calls for private equity and real estate funds, the need for immediate cash was even greater. In some cases, the universities cut spending, halting building programs and even eliminating some faculty and staff positions, while raising tuition at higher rates than in prior years.

When cash is scarce, it can be difficult to have such large allocations to illiquid alternative investments and such small allocations to cash and fixed income. Sheikh and Sun (2012) explain that the cash and fixed-income holdings of an endowment should be at least 6% to 14% of assets to avoid liquidity crises in 95% of market conditions. To completely eliminate liquidity risk, cash and fixed-income holdings may need to be as high as 35%, far above the allocations that most endowments are comfortable making, given their high expected return targets. By drawing down this cash cushion, the endowment can continue to fund spending to support the university budget while avoiding a liquidity crisis that would lead to the distressed sale of assets at the low point in the market or an emergency increase in the debt burden. Greater cash holdings are necessary for universities with larger outstanding commitments to private equity and real estate funds, greater leverage, higher spending rates, more frequent rebalancing, or larger allocations to less liquid assets.

To avoid liquidity crises, Siegel (2008) suggests laddering allocations to private equity and real estate funds, ideally at a schedule in which distributions from maturing funds are sufficient to fund capital calls of partnerships of more recent vintages. When adding real estate and private equity partnerships to the portfolio, investors are encouraged to spread the new commitments over multiple years rather than making a large initial commitment in a single vintage year. In addition to spreading capital commitments over time, Siegel suggests that liquidity can be improved by growing the gift income of the endowment, borrowing, or reducing the allocation to less liquid alternative investments. Private equity and real estate partnerships are less liquid investments, while commodity futures funds and hedge funds with lockups of one year or less are more liquid alternative investments.

Leverage can also create liquidity issues. Short-term leverage, such as that provided by prime brokers to hedge funds, may not be sustainable or affordable during times of crisis. When credit lines are reduced or not renewed, investors may have to repay loans on short notice, which can require the sale of investments at very low prices. Many fixed-income arbitrage and convertible bond arbitrage funds suffered significant losses during the most recent crisis, as a reduction in leverage from eight times to four times required the immediate sale of half of the portfolio. When the market knows that these sales are coming, and a number of hedge funds are simultaneously forced to sell due to credit line reductions as well as investor redemptions, liquidity risk is extreme as buyers of these fixed-income assets wait to purchase until the prices of the convertible and mortgage-backed bonds have fallen precipitously. Endowments and foundations that invest in leveraged hedge funds must be prepared for the potentially large drawdowns in these strategies, as well as the potential for the erection of gates that prevent investors from redeeming their assets from hedge funds during times of market crisis.

The liquidity crisis of 2008 brought criticism to the endowment model. Williamson (2011) quotes Daniel Wallick, principal of Vanguard's Investment Strategy Group, as saying that the endowment world's pre-2008 blind emulation of the Yale approach has passed. Endowments and foundations today need to focus on having greater access to liquidity in their funds, which may lead both to declines in the commitments toward future private equity and real estate funds, and to increases in the cash and fixed-income allocations. Within alternatives, the focus has turned toward more liquid holdings, such as equity hedge funds and commodity futures investments. Between June 2008 and June 2010, Commonfund notes that cash allocations increased by 4%, while U.S. equity allocations fell by 8% and international equity allocations declined by 2%. Alternative investments, especially those with greater liquidity and lower volatility, increased allocations by 6%.

Not everyone, though, thinks that the endowment model has passed its prime. Keating (2011) believes that after some tweaks in liquidity, conviction in the endowment model has actually strengthened. He notes that the Harvard University endowment has changed its cash target from –5% to +2%, while reducing its uncalled capital commitments to real estate and private equity partnerships by more than $4 billion in the past two years. Similarly, Yale University increased its cash holdings to 4%, while putting external lines of credit into place. Keating (2010) states that the liquidity crisis was not caused by an overallocation to alternative investments but by an underallocation to fixed-income and cash investments.

There are important lessons to be learned from the experiences of pension funds and endowments during the most recent financial crisis. Plan sponsors, portfolio managers, and asset allocators could use the framework set forth by the Bank for International Settlements to create a robust process to estimate their liquidity needs and establish a clear liquidity risk tolerance that reflects the needs of their current and future beneficiaries. They should establish sound processes for identifying, measuring, monitoring, and controlling liquidity risk. This process should include estimates of future cash flows arising from both assets and liabilities. A sound and robust risk management process should allow pension funds and endowments to take full advantage of the available investment opportunities, including earning premiums for bearing liquidity risk at levels their institutions can tolerate.

3.5.3 Rebalancing and Tactical Asset Allocation

Among large endowments, the growth rate of allocations to alternative investments may be approaching the largest possible level. Other institutional investors continue to increase allocations to alternative investments in hopes of catching up with the top universities in terms of both returns and the size of the assets allocated to alternative investments. In addition to a large allocation to alternative investments, emulating the largest endowments also requires aggressive rebalancing, careful sourcing of top-performing managers, and embracing liquidity risk. This is easier said than done, however, as inevitable market crises will test the patience and liquidity structures of investors with large holdings in alternative investments.

Another reason to maintain liquidity in an endowment or a foundation portfolio is to facilitate rebalancing activity. Swensen (2009) believes strongly in keeping portfolio weights close to the long-term strategic weights, a practice that requires regular rebalancing. Without rebalancing, the asset allocation of the portfolio will drift, with the asset classes earning the highest returns rising in weight relative to the rest of the portfolio. Assuming that the highest-performing asset class is also more volatile and increasingly overvalued, the risk of the portfolio rises significantly when rebalancing activity is delayed. Market price action makes it relatively easy to rebalance publicly traded securities, as the investor is buying as prices fall and selling as prices rise. Investors who rebalance are providing liquidity to the market, and liquidity providers often get paid for providing that service to other investors. This is the time when value is created, as many times purchases made during a time of price weakness can create significant value. It can take courage, though, to buy an underweighted asset class when prices are falling and most other investors are selling. To the extent that bonds increase in value as a flight-to-quality asset when equities decline, investors may need to move quickly to rebalance before returns start to move in the opposite direction.

Rebalancing, however, can be regularly undertaken only in liquid asset classes. Within alternatives, hedge funds may have quarterly redemption windows and lockup periods of one to three years. Private equity and real estate funds must typically be held until assets are fully distributed, a process that can take 10 to 12 years. Funding capital calls to private equity and real estate funds can change the asset mix, as traditional investments are typically sold to fund the increasing allocation to the less liquid alternative investments. To the extent that alternative investments have net asset values that are smoothed or reported with a time lag, publicly traded investments will decline in allocation rapidly during times of crisis. It is important to understand the role of pricing in these less liquid asset classes, as the net asset value adjusts slowly to changes in public market valuation. Investors may react by rebalancing only within the liquid alternatives and traditional assets, while slowly changing allocations to less liquid alternative investments by modifying the size of future commitments.

There are a number of approaches to rebalancing, such as those discussed by Kochard and Rittereiser (2008). Some investors will rebalance on a calendar basis, for example, after discussions at a quarterly meeting of the investment committee. Other investors will tie the rebalancing activity to the actual asset allocation when compared to the long-term policy asset allocation. While some investors have exact targets for the domestic equity allocation, such as 30%, others might have ranges of 25% to 30%. Those with an exact target may establish a rebalancing deviation, such as a decision to rebalance when the equity allocation has strayed 2% from its target weight. Investors with asset allocation ranges may wait to rebalance until the allocation has moved outside the range. When range-based investors rebalance, they must also decide whether to rebalance to the closest edge of the range or to the center of the range.

For liquid investments, rebalancing can be accomplished through the use of securities or derivatives. Investors seeking to rebalance during late 2008 or early 2009 needed to sell fixed income and buy equity securities in order to restore the liquid portion of the portfolio back to the strategic asset allocation weights. While the crisis led to both declines in equity prices and increases in yields on risky fixed-income securities, the drawdown in the equity portfolio was much larger. As spreads on investment-grade and high-yield corporate bonds widened significantly, sovereign bond yields declined due to the flight-to-quality response. Even though investors desired to rebalance, many managers of fixed-income funds, especially in convertible bonds or mortgage-backed securities, had restricted liquidity by suspending redemptions or implementing gates. Experienced investors noticed a tremendous opportunity to rebalance using the derivatives markets. When the S&P 500 Index traded above 1,400 in May 2008, the 10-year Treasury yielded 3.8%. At the market low in March 2009, Treasury notes had rallied to a yield of 2.8%, while the S&P 500 traded below 700. There was quite a window for rebalancing, as the S&P 500 was valued at below 900 from the end of November 2008 to the end of April 2009. Investors who sold 10-year Treasury note futures and bought futures on the S&P 500 at any time during late 2008 or early 2009 had a tremendous profit from the rebalancing trade. This was because by the end of 2009, Treasury yields had returned to 3.8% while the S&P 500 had moved above 1,100, producing a profit of at least 24% on the equity trade alone. Investors who kept their fixed-income funds intact while hedging the change in Treasury yields multiplied their profits as yield spreads declined from record levels in the spring to more normal levels by year-end.

Those schooled in options theory may notice that rebalancing activity is simply a short strangle trade, where both out-of-the-money calls and puts are sold. If the investor is committed to reducing the equity allocation after prices have risen 10%, it can make sense to sell index call options 10% above the market. This brings discipline to the rebalancing process and allows the fund to earn income through the sale of options premium. This income can be either spent by the sponsor of the endowment or foundation fund or used to reduce the risk of the investment portfolio. Similarly, committing to buy equities after a 10% decline could be implemented through the sale of equity index put options with a strike price 10% below the current market level. The simultaneous sale of calls and puts at the same strike price is termed *selling straddles*, while selling out-of-the-money calls and puts at different strike prices is termed *selling strangles*. While this approach can earn significant options premium and bring discipline to the rebalancing process, it is not without risk. The greatest risk is when the market makes a move larger than 10% in either direction. The sale of options guarantees that rebalancing will occur at the level of the strike price, while those without options hedges may be able to rebalance after the market has moved by 20% to 30%. Investors can reduce the risk of using options to rebalance by selling call spreads and put spreads rather than selling strangles. While the purchase of further out-of-the-money options reduces risk and opportunity costs, the net premium earned from the sale of spreads will be less than that earned for selling strangles. Of course, there can be significant fear or euphoria after such a move, and some managers may hesitate to rebalance due to the foibles understood by students of behavioral finance.

Some endowments may employ internal tactical asset allocation (TAA) models or external asset managers offering TAA strategies. As opposed to strategic asset allocation, which regularly rebalances back to the long-term target weights, tactical asset allocation intentionally deviates from target weights in an attempt to earn excess returns or reduce portfolio risk. TAA models take a shorter-term view on asset classes, overweighting undervalued assets and underweighting overvalued assets. While the risk and return estimates underlying the strategic asset allocation are typically calculated for a 10- to 20-year period, the risk and return estimates used by tactical asset allocation are typically much shorter, often between one quarter and one year.

Tactical models are most useful when markets are far from equilibrium, such as when stocks are expensive at 40 times earnings or when high-yield bond spreads are cheap at 8% over sovereign debt. TAA models can employ valuation data, fundamental and macroeconomic data, price momentum data, or any combination of the three.

A number of alternative investment styles employ TAA analysis. Managed futures funds focus on price momentum, while global macro funds more commonly analyze governmental actions to predict moves in fixed-income and currency markets. TAA funds may employ both methodologies but are different from managed futures and macro funds. First, managed futures and macro funds take both long and short positions and often employ leverage; TAA funds are typically long-only, unlevered funds. Second, TAA funds may reallocate assets across a small number of macro markets, whereas managed futures and global macro funds may have a much larger universe of potential investments.

Because TAA strategies can be difficult to employ successfully, many investors will place limits on the size of tactical positions. For example, when stocks are overvalued, the equity allocation may be 10% below the long-term target weight. If the fund were allowed to swing between 100% equity and 100% fixed-income allocations, substantial opportunity costs could be incurred. In this case, a TAA portfolio may have been fully invested in fixed income when equity prices moved 25% higher between the end of February and May 2009. Due to the similarity in long-term return estimates, it is less risky to tactically allocate between assets of similar risk and return (hedge funds versus commodities versus stocks) than between assets of different risk and return (equities versus cash), which carry a much higher opportunity cost.

3.5.4 Tail Risk

In the foreword to Swensen (2009), Charles Ellis comments that Yale was good at playing defense, because the endowment was built to withstand the inevitable storms that face capital markets. This resiliency was put to a severe test, as the Yale endowment lost 25% of its value in the 12 months ending June 2009. This has been termed a *tail event*, in that the returns were at the extreme left tail of the endowment's return distribution.

Bhansali (2008, 2010, 2011) has repeatedly encouraged investors to manage the risk of catastrophic loss of portfolio value, termed *tail risk*. When portfolios preserve value during bear markets, the long-term value of the endowment fund can be increased. The key to minimizing drawdowns is to build some protection into the portfolio by making an allocation to assets that will maintain value or even rise in value during times of crisis.

The most straightforward hedge is an increased weight on cash and risk-free debt in the portfolio. A rising allocation to cash, however, will reduce the expected return of the portfolio and potentially lead to lower long-term wealth. The most aggressive endowment and foundation investors have clearly not used cash and fixed income as a tail hedge, as the allocation to this defensive asset class is typically quite low.

In times of market stress, correlations between many types of assets tend to rise. This increases portfolio volatility above that assumed in the mean-variance optimization that may have been used to determine the initial asset allocation. Page, Simonia, and He (2011) state that private equity, real estate, and hedge funds earn a liquidity premium but sell put options on liquidity risk. When equity markets decline in a crisis, these alternative investments also experience losses. Investors may wish to estimate the equity betas of their portfolios during periods of both normal and extreme market moves. After determining the equity exposure of the portfolio, including other assets that behave like equity during a time of crisis, the hedging process can begin. Page, Simonia, and He estimate that the typical endowment portfolio may derive over 70% of its risk from equity markets. With an allocation of 31% equity, 17% fixed income and cash, and 52% alternative investments, the risk can be parsed two ways. First, the risk can be decomposed into 81% equity and 19% corporate spreads, currency, commodity, and other. When including liquidity risk, this risk decomposition changes to 61% equity; 25% liquidity; and 14% corporate spreads, currency, commodity, and other.

A second method to reduce tail risk is to employ options hedges on the equity-linked portion of the portfolio. **Equity options hedges** are positions established in equity options for the primary purpose of reducing the equity risk of a portfolio, such as the purchase of a put option. Equity put options provide the purest hedge against tail risk, offering the potential to provide a greater than 500% return during times of increasing systemic risk. For example, an investor who spends 5% of portfolio value each year on equity put options may expect those options to be worth 25% of the portfolio value at the bottom of the bear market. However, the simple purchase of equity put options can be quite expensive. Further, this approach may simply be smoothing returns: transferring the losses on options in good years to profits on options in years of declining equity markets.

The cost of equity options hedges can be reduced through the use of collars or put spreads. In a collar, a call option is sold above the market. While this limits the potential return from the equity-linked portion of the portfolio, the premium earned from the sale of the call option can offset the cost of the put options. In a put spread, the investor purchases one put option at perhaps 10% out-of-the-money, while selling a second put option at perhaps 25% out-of-the-money. This strategy can insure losses on the equity portfolio of up to 15%, but after the market has fallen 25%, the investor participates fully in market declines. The cost of a put spread may be 30% to 70% less than the cost of a long put option, depending on the implied volatility, strike price, and maturity of each option.

Bhansali (2008, 2010, 2011) advocates an opportunistic approach to hedging, which leans heavily on the idea that correlations between risky assets rise in a crisis. While equity options are a pure hedge against the dominant risk in most investor portfolios, they are often the most expensive. During extreme market events, markets with cheaper hedging vehicles may have moves as large as equity markets. Bhansali proposes building portfolios of put option hedges from the currency, commodity, and credit markets, and call options on volatility indices, buying hedges when they are cheap and selling hedges when they are expensive. To take advantage of the flight-to-quality nature of market crises, call options on high-quality bonds may also be employed.

To the extent that this basket hedging approach is exchange traded, the hedges will be liquid even during a crisis. Investors can sell their hedges immediately after a large market decline, gaining access to cash at a time when asset markets are experiencing declining liquidity and new borrowings may be expensive or infeasible. This defensive strategy may allow investors to play offense by buying assets from other investors who are in need of cash. During 2008 and 2009, investors with cash were able to buy shares of hedge fund, private equity, and real estate partnerships at a discount in the secondary market from investors who were unable to raise the cash to fund their spending rates or capital calls.

Investors need to be careful, though, when hedges are purchased in the over-the-counter market. Trades in the over-the-counter market incur counterparty risk, which can be at its highest point during times of market crisis.

Within each asset class, investors can structure allocations to reduce exposure to extreme market events. Just as Swensen does at Yale, the fixed-income portfolio can focus on high-quality bonds, which will grow in value during a crisis, while avoiding corporate bonds, whose yield spreads widen quickly during a market stress event. With hedge funds, it may be wise to reduce allocations to arbitrage strategies, such as convertible arbitrage or mortgage-backed securities arbitrage, which rely on tightening spreads, the availability of leverage, and liquid markets to earn their returns. Some hedge fund strategies have historically risen in value during times of market crisis. Macro, managed futures, and some volatility arbitrage funds are designed to have their largest returns during times of extreme market moves, so some investors specifically allocate assets to these strategies to reduce the tail risk of their portfolios. While these hedge fund strategies are not as certain to perform as a put options strategy during times of market stress, the expected cost of these strategies is lower, as their long-term return far exceeds the negative expected return of programs that regularly purchase equity put options.

3.6 CONCLUSION

It is important for investors and analysts to understand the endowment model, which seeks a high allocation to alternative investments in order to meet aggressive targets requiring long-term returns that exceed the rate of inflation by at least the amount of required spending. Portfolio managers and asset allocators seeking to emulate the returns of the most successful endowments must realize that simply mimicking the asset allocations of top endowments will not guarantee similar returns, as endowments have historically added significant value through manager selection and market timing.

Those wishing to replicate the results of the most successful endowment and foundation investors need to consider the risks to inflation, liquidity, and extreme market events, while adding value through rebalancing and selecting active managers. A focus on alternative investments also requires a greater degree of investment manager due diligence, evaluating both investment and operational risks.

NOTE

1. SAT is a registered trademark of the College Board.

REFERENCES

Alliance Bernstein. 2010. "Deflating Inflation: Redefining the Inflation-Resistant Portfolio." April.

Ang, A., and K. N. Kjaer. 2011. "Investing for the Long Run." November. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1958258.

Anson, M. J. P. 2010. "Measuring a Premium for Liquidity Risk." *Journal of Private Equity* 13 (2): 6–16.

Aragon, G. 2004. "Share Restrictions and Asset Pricing: Evidence from the Hedge Fund Industry." *Journal of Financial Economics* 83:33–58.

Barber, B. M., and G. Wang. 2013. "Do (Some) University Endowments Earn Alpha?" *Financial Analysts Journal* 69 (5): 26–44.

Bary, A. 2009. "Big Squeeze on Ivy League Endowments." *Barron's*, July 1.

Bhansali, V. 2008. "Tail Risk Management." *Journal of Portfolio Management* 34:68–75.

———. 2010. "Vineer Bhansali Discusses PIMCO's Approach to Tail Risk Hedging." PIMCO. http://europe.pimco.com/EN/Insights/Pages/Bhansali%20QA%20Tail%20Risk%20-%20Feb%202010.aspx.

———. 2011. "Cash vs. Tail Risk Hedging: Which Is Better?" PIMCO. http://europe.pimco.com/EN/Insights/Pages/Cash-vs-Tail-Risk-Hedging-Which-Is-Better.aspx.

Brinson, G. P., L. R. Hood, and G. L. Beebower. 1986. "Determinants of Portfolio Performance." *Financial Analysts Journal* 42 (4): 39–44.

Brinson, G. P., B. D. Singer, and G. L. Beebower. 1991. "Determinants of Portfolio Performance II: An Update." *Financial Analysts Journal* 47 (3): 40–48.

Brown, K., L. Garlappi, and C. Tiu. 2010. "Asset Allocation and Portfolio Performance: Evidence from University Endowment Funds." *Journal of Financial Markets* 13 (2): 268–94.

Chacko, G., C. L. Evans, H. Gunawan, and A. L. Sjoman. 2011. *The Global Economic System: How Liquidity Shocks Affect Financial Institutions and Lead to Economic Crises*. Upper Saddle River, NJ: FT Press.

Commonfund Institute. 2013. "Outsourced Investment Management: An Overview for Institutional Decision Makers." November.

Ford Foundation. 1969. *Managing Educational Endowments: Report to the Ford Foundation*. New York: Ford Foundation.

Ibbotson, R. G., and P. D. Kaplan. 2000. "Does Asset Allocation Policy Explain 40, 90, or 100 Percent of Performance?" *Financial Analysts Journal* 56 (1): 26–33.

Keating, T. 2010. "The Yale Endowment Model of Investing Is Not Dead." RIABiz, April 20.

———. 2011. "How the Harvard and Yale Endowment Models Changed to Avoid a Repeat of 2009." RIABiz, February 15.

Khandani, A., and A. W. Lo. 2011. "Illiquidity Premia in Asset Returns: An Empirical Analysis of Hedge Funds, Mutual Funds, and U.S. Equity Portfolios." *Quarterly Journal of Finance* 1 (2): 205–64.

Kochard, L. E., and C. M. Rittereiser. 2008. *Foundation and Endowment Investing: Philosophies and Strategies of Top Investors and Institutions*. Hoboken, NJ: John Wiley & Sons.

Lerner, J., A. Schoar, and J. Wang. 2008. "Secrets of the Academy: The Drivers of University Endowment Success." *Journal of Economic Perspectives* 22:207–22.

Lerner, J., A. Schoar, and W. Wongsunwai. 2007. "Smart Institutions, Foolish Choices? The Limited Partner Performance Puzzle." *Journal of Finance* 62 (2): 731–64.

Li, H., X. Zhang, and R. Zhao. 2011. "Investing in Talents: Manager Characteristics and Hedge Fund Performances." *Journal of Financial and Quantitative Analysis* 46 (2): 59–82.

Lord, M. 2014. "University Endowment Committees: How a Learning Orientation and Knowledge Factors Contribute to Portfolio Diversification and Performance." *European Journal of Finance*. doi:10.1080/1351847X.2013.879536.

Mladina, P., and J. Coyle. 2010. "Yale's Endowment Returns: Manager Skill or Risk Exposure?" *Journal of Wealth Management* 13 (2): 43–50.

Page, S., J. Simonia, and F. He. 2011. "Asset Allocation: Systemic Liquidity as a Risk Factor." *Trading* 1:19–23.

Sadkay, R. 2009. "Liquidity Risk and the Cross-Section of Hedge-Fund Returns." Working paper, Boston College.

Sheikh, A., and J. Sun. 2012. "Defending the 'Endowment Model': Quantifying Liquidity Risk in a Post–Credit Crisis World." *Journal of Alternative Investments* 14, no. 4 (Spring): 9–24.

Siegel, L. 2008. "Alternatives and Liquidity: Will Spending and Capital Calls Eat Your 'Modern' Portfolio?" *Journal of Portfolio Management* 35 (2): 103–14.

Swensen, D. 2009. *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*. New York: Simon & Schuster.

Takahashi, D., and A. Alexander. 2002. "Illiquid Alternative Asset Fund Modeling." *Journal of Portfolio Management* 28 (2): 90–100.

Williamson, C. 2011. "Endowment Execs Focus on Liquidity, Volatility in Post-Crisis Market." *Pensions & Investments*, March 7.

———. 2013. "Outsourced Assets Catapult 59% to $1 Trillion in Two Years." *Pensions & Investments*, July 8.

CHAPTER 4

Pension Fund Portfolio Management

Pension plans (also known as pension schemes or superannuation plans) manage assets that are used to provide workers with a flow of income during their retirement years. This chapter provides an overview of pension funds as they relate to alternative investing.

4.1 DEVELOPMENT, MOTIVATIONS, AND TYPES OF PENSION PLANS

Because pension schemes may control the largest pool of capital in the world, asset managers need to be aware of the goals and challenges of managing these plans.

4.1.1 Development of Pension Plans

In a study of 19 developed countries, private and public pension plan assets totaled over $35 trillion, averaging 80% of gross domestic product (GDP) (Willis Towers Watson 2016). At the end of 2013, World Bank data[1] shows that over 86% of the workforce is covered by some form of pension or government retirement plan in 22 of 23 high-income countries. The median country in Europe and Central Asia covers 67.9% of the workforce, while coverage in the rest of the world, especially Africa, is very low, with a median coverage of just 20.3% of workers. The world's top 15 pension plans controlled over $5,397 billion in assets in 2015 (see Exhibit 4.1).

In most of the developed world (North America, Europe, Japan, and Australia), life expectancy exceeds 80 years. Workers may start a career around age 20, work for approximately 40 years, and retire from work between ages 60 and 67. Workers need to save during their careers to maintain an adequate standard of living during retirement. It can be difficult for an individual worker to adequately plan for retirement, as investment returns and one's life expectancy are unknown. Depending on their chosen career and income, workers may lack either the ability to save or the investment knowledge to appropriately invest their assets.

4.1.2 Motivations to Using Pension Plans

There are a number of reasons why pension plans can be attractive, both for employers and for employees. Companies offering pension plans may be able to attract and retain higher-quality employees while employees may seek out companies offering

EXHIBIT 4.1 The World's Largest Pension Plan Sponsors, 2015

Fund	Country	Assets ($ million)
Government Pension Investment	Japan	$1,143,838
Government Pension Fund	Norway	$ 884,031
National Pension	South Korea	$ 429,794
Federal Retirement Thrift	U.S.	$ 422,200
ABP	Netherlands	$ 418,745
California Public Employees	U.S.	$ 296,744
National Social Security	China	$ 247,361
Canada Pension*	Canada	$ 228,431
PFZW*	Netherlands	$ 215,006
Central Provident Fund	Singapore	$ 207,872
Local Government Officials	Japan	$ 194,696
California State Teachers	U.S.	$ 186,954
Employees Provident Fund	Malaysia	$ 184,697
New York State Common	U.S.	$ 178,252
New York City Retirement	U.S.	$ 158,702

*As of March 31, 2015.
Source: Pensions & Investments.

strong pension benefits. Employees value the income promised by a pension plan, which may be used as a substitute for their personal savings. In many countries, retirement plan assets grow on a tax-deferred basis. Employees' and employers' contributions to retirement plans are not taxed in the year that the contributions are made. The gains on the investment portfolio are not taxed in the year they are earned, but taxes are paid by employees when the assets are withdrawn during retirement. Ideally, the employee will pay a lower tax rate during retirement than during the working years, which further increases the tax benefit of pension plan investments. Finally, there are economies of scale with large pension plans. These large plans are able to meet substantial investment minimums required of many alternative investments, and the pooling of mortality risk allows the pension plan to allocate a larger portion of assets to illiquid assets to earn a liquidity premium.

In contrast to what occurs when employees individually save for retirement, pension funds have several advantages. First, the pension fund can hire internal staff and external managers who are highly trained in finance to watch the investment portfolio on a daily basis. Economies of scale are also earned by large pension plans, as larger investment sizes can reduce investment fees and afford a larger staff.

Pension plans can also make long-term investments, with a time horizon that may be as long as the lifetime of the youngest employee. Asset allocation decisions are made with the average employee in mind. When individual investors make retirement investments, asset allocation becomes inherently more conservative over time, as the employee's lifetime is uncertain and the ability to fund investment losses during retirement is limited. Mortality risk, the risk that someone will die earlier than expected, is highly uncertain for an individual investor, but can be quite predictable when averaged over a large number of employees and retirees covered by a pension plan. Longer lifetimes require larger retirement assets. For an individual investor, spending rates may be conservative, again because the life span is uncertain. However, for a pension plan with known benefits, the asset allocation and benefit levels may not be significantly impacted by the death of a single beneficiary. Longevity risk, the risk that individuals will live longer than anticipated and outlive their resources, affects different investors in different ways. For life insurance companies, the risk is that beneficiaries die at a younger age than predicted, as the life insurance benefit will be paid at an earlier date and a higher present value. For individuals and pension plans, the risk is that lifetimes will be longer than anticipated, as retirement spending or retirement benefits will last for a longer period, requiring a larger number of monthly benefit payments or months of retirement spending. Life insurance companies have an interest in selling hedges for longevity risk, as this is a natural fit for their life insurance book. In life insurance, benefit payments rise when the insured dies young, while in longevity risk (or annuities) benefit payments rise as the insured has a longer than expected life.

4.1.3 Three Basic Types of Pension Plans

There are three basic types of pension plans: defined benefit, governmental social security plans, and defined contribution. Each plan varies in the asset management risks and rewards, and whether the employer, the employee, or taxpayers have the ultimate risk for the performance of the investment portfolio.

There are hybrid types of schemes such as cash balance plans that share features with both defined benefit and defined contribution plans. A **cash balance plan** is basically a defined benefit plan, where the pension benefits are maintained in individual record-keeping accounts that show the participant the current value of his or her accrued benefit and facilitate portability to a new plan.

4.2 RISK TOLERANCE AND ASSET ALLOCATION

Chapters 1 and 2 discussed the asset allocation process and techniques for determining optimal asset allocation. It was explained that the first step in the asset allocation process is to identify the asset owners' objectives and constraints. This information is then used in the second step to create an investment policy statement, which will be used by the portfolio manager to implement asset allocation strategies that are consistent with the asset owner's needs.

4.2.1 Three Approaches to Managing the Assets of Defined Benefit Plan

In managing a defined benefit (DB) plan, the primary objective is to fund the plan's liabilities. In creating optimal portfolios, one needs to develop an appropriate measure of risk for the fund and account for factors that affect the plan sponsor's level of risk tolerance.

The risk of a DB plan can be measured from three difference perspectives: assets, assets and liabilities, and integrated. Different risk metrics such as volatility, value at risk, conditional value at risk, and risk budgeting can be employed for measuring risk. Here, volatility is used as the risk metric.

4.2.1.1 Asset-Focused Risk Management When the risk of a DB plan is measured using assets only, the portfolio manager could consider the volatility of the rate of return on the plan's assets as a measure of its riskiness. In this context, cash and cash equivalents are considered riskless. Subject to the plan sponsor's degree of risk tolerance and other constraints (e.g., liquidity needs), optimal allocations are created so that the portfolio can earn the highest possible return.

4.2.1.2 Asset-Liability Risk Management Perspective From an asset-liability perspective, the risk of the DB plan is measured in terms of the volatility of its surplus. For example, suppose the plan's assets and liabilities at time t are given by A_t and L_i, respectively. Then the risk from an asset-liability framework can be measured using the standard deviation of $(A_t - L_t)$. Everything else being the same, the standard deviation is lower as the correlation between assets and liabilities increases. Therefore, in this context, cash or cash equivalents are not riskless assets. An asset that is volatile but positively correlated with changes in liabilities is considered to have a lower risk.

4.2.1.3 Integrated Asset-Liability Risk Management In an integrated approach, the plan's funding status is integrated with the plan sponsor's operations. In this case, it is recognized that future shortfalls in the plan's funding status are likely to require larger contributions by the plan sponsor. Therefore, everything else being the same, it will be desirable if there is a negative correlation between $(A_t - L_t)$ and profitability of the plan sponsor. As a result, whenever there is a decline in $(A_t - L_t)$, there is likely to be an increase the firm's profitability, making it easier for the firm to contribute to the plan. In this context, an asset that is positively correlated with liabilities may be considered rather risky if it is positively correlated with the firm's operational strength as well.

There are other ways of measuring the riskiness of a pension fund in an integrated approach. For example, the plan sponsor may wish to reduce the volatility of the firm's equity. Equity is given by:

$$E_t = OA_t - OL_t + (A_t - L_t) \tag{4.1}$$

where E_t is the market value of equity, and OA_t and OL_t are operating assets and liabilities, respectively. In this context, the volatility of equity will depend on the correlations between the changes in surplus, operating assets and operating liabilities. In particular, everything else being the same, assets that have low correlations with operating assets and high correlations with operating liabilities are considered less risky.

4.2.2 Four Factors Driving the Impact of Liabilities on a Plan's Risk

The primary liabilities of a pension plan are the benefits due to participants. We saw that unless the fund measures risk from an assets-only perspective, changes in the plan's liabilities affect the plan's measure of riskiness. Four factors impact the value of liabilities the most: interest rates, inflation, retirement cycle, and mortality. Changes in interest rates are the most important factor, as they affect the discount rate used to calculate the present value of future obligations. A decline in interest rates will increase the current value of future liabilities. Inflation affects the future value of liabilities to the degree that benefits are directly or indirectly affected by inflation. For instance, the benefits are typically tied to salary, and salaries tend to increase with inflation. Also, in some cases, benefits might be tied to inflation directly through cost of living adjustments.

The retirement cycle is the third factor, which tends to be highly predictable. Given the age profile of employees and the firm's retirement policy, it will not be difficult to forecast the number of employees who are likely to retire in future years. However, changes in the economic environment could create uncertainty regarding the retirement process. For instance, the number of employees and their tenure at the firm may change over time, either if the firm reduces staff in a recession or if employees leave during times of economic growth to seek other employment. The final factor affecting liabilities is the change in the mortality rate. While longevity risk (i.e., a declining mortality rate) can pose a serious threat to the financial viability of a DB plan, the risk is highly predictable and has a small influence on the value of the liabilities in the short run.

4.2.3 Five Major Factors Affecting the Risk Tolerance of the Plan Sponsor

Having discussed various ways of measuring the riskiness of a fund, the next step is to examine factors that affect the risk tolerance of the plan sponsor. Five factors have the most impact on a plan sponsor's risk tolerance: the funding status of the plan, the size of the fund, expected size of future contributions relative to the employer's cash flow, the employer's financial position, and the employees' characteristics.

The DB plan can be overfunded, underfunded, or fully funded. In general, the larger the deficit, the less tolerance for risk the plan sponsor is likely to have. It is important to note that this does not mean that an underfunded plan will always be less risky. The amount of risk assumed by the plan is determined by the interaction of the sponsor's tolerance for risk and its objective. For instance, consider a slightly underfunded plan, where the plan sponsor has a low tolerance for risk. The sponsor will have to assume some risk if the objective is to earn a higher rate of return than the growth rate of liabilities to improve the plan's funded status. In addition to the funding status of the plan, the size of the plan influences the sponsor's risk tolerance. The risk tolerance of the plan sponsor rises with the size of the plan's liabilities relative to the size of the sponsor's assets.

The third factor that influences the sponsor's risk tolerance is the potential size of future contributions relative to its projected free cash flows. If the sponsor is expected to generate significant free cash flows in future, then its capacity to take risk increases. The sponsor should evaluate the projected annual contributions that will be needed to cover current and future obligations. These potential contributions must then be compared to the firm's future free cash flows. If the future free cash flows are relatively small, then the firm's tolerance for risk is reduced.

The fourth factor is the sponsor's financial position, including the potential free cash flows expected to be generated by the firm. Plan sponsors with a high debt-to-equity ratio will have a lower tolerance for risk. Not only are these organizations already exposed to business fluctuations, but they will not be able to issue debt to contribute to the fund. Employers that have little debt on their balance sheets may find it advantageous to issue debt to contribute to the fund.

The final factor affecting the sponsor's risk tolerance is the demographics of the employees. In general, sponsors that have younger employees tend to have more tolerance for risk. The primary reason for this is that that there is more time to reduce any potential shortfall in the fund's assets. For instance, a plan that is modestly underfunded could become fully funded after a long period even if the plan's assets outperform its liabilities by only a few basis points each year. Therefore, the sponsor will have a higher capacity to assume some surplus risk. On the other hand, an employer with an aging workforce will have a smaller capacity for risk tolerance. Also, an aging workforce increases the sponsor's needs to have enough liquidity in the fund, which reduces the tolerance for liquidity risk.

4.2.4 Strategic Asset Allocation of a Pension Plan Using Two Buckets

As discussed in Chapter 1, the next step after evaluating the asset owner's objectives is to develop an investment policy statement. In an asset-liability framework, the focus will be on the relationships between asset and liabilities. A simple approach to creating a strategic asset allocation for a DB plan is to consider the entire portfolio consisting of two separate buckets: a hedging bucket and a growth bucket.[2]

The hedging portfolio is created to mimic the growth of the liabilities. The goal is to reduce the volatility of the fund's surplus, whether it is positive or negative. Given the previous discussion, this portfolio is likely to consist of allocations to asset classes whose returns are negatively correlated with changes in interest rates, positively correlated with the inflation rate, and positively correlated with increased longevity in the population. For instance, a combination of long-term bonds, inflation-indexed bonds, health care stocks, or longevity-related derivatives could be used to create the hedging bucket. In addition, the investment policy statement may consider the correlation between the hedging portfolio and the sponsor's operating revenues. In this case, the hedging portfolio's return should not be highly correlated with changes in the sponsor's future free cash flows. Three approaches may be used to create the hedging bucket in the asset-liability framework: duration matching, cash flow matching, and overlay.

In the **duration matching approach**, the duration of the hedging bucket matches the duration of the liabilities. The approach is simple to implement but requires careful monitoring as changes in the yield curve and credit spreads will impact the duration of assets and therefore the portfolio will have to be rebalanced.

In the **cash flow matching approach**, the hedging portfolio is constructed such that its estimated future cash inflows match the expected outflows associated with liabilities at each prospective point in time. In this case, a customized portfolio consisting of zero-coupon bonds that mature at the time of all future payments in amounts equal to those payments is constructed. Compared to duration matching, this approach is harder to implement because appropriate instruments may not be available. For example, very long-term zero-coupon bonds may not be available and may be costly to create using strips, which are principal-only securities created when the interest and principal of a bond are separated or stripped.

In the **overlay approach**, the plan sponsor employs derivatives to create a hedging bucket. It is important to note that this could lead to leveraged positions, which will increase the overall risk of the portfolio. The advantage of this approach is that the sponsor might be able to create the hedging bucket without the need to sell portions of the growth bucket. For instance, the fund can enter into an interest rate swap to receive the fixed rate and pay the variable rate. If interest rates decline, the portfolio will benefit from the swap position as the value of liabilities increases. At the same time, the sponsor can maintain its allocation to the growth bucket, which is expected to outperform the short-term rate of interest.

The growth bucket is constructed with the expectation that it will outperform the liabilities and therefore reduce the sponsor's future contributions to the fund. The size of the allocation to the growth bucket depends on the capacity of the sponsor to assume some surplus risk. As the sponsor's tolerance for risk increases, two courses of action will become open. First, a larger portion of the overall assets can be allocated to the growth bucket. In this case, the riskiness of the growth bucket is held constant, but a larger portion of the portfolio is allocated to it. Second, the sponsor can decide to keep the relative sizes of hedging and growth buckets unchanged, but increase the growth bucket's allocations to less liquid and riskier asset classes in order to earn a higher rate of return. For instance, the sponsor may decide to increase allocations to private equity, timber, private real estate, and hedge funds with long lockups.

4.3 DEFINED BENEFIT PLANS

Defined benefit (DB) plans provide a guaranteed income to retirees, but can be risky for employers. In a **defined benefit plan**, the employer takes all of the investment risk while offering a guaranteed, formulaic benefit to retirees.

For example, consider an employer that offers a retirement benefit of 1.5% of salary for each year the employee worked before retirement. Typically the salary used to calculate the benefit is based on a simple formula such as the average annual salary over the final three years of employment. If the salary to which the benefits apply is \$50,000 and the employee has worked for 40 years, the retiree will be paid retirement benefits in the amount of \$30,000 per year (1.5% × 40 years × \$50,000) for the rest of the retiree's life. This provides the worker with a **retirement income-replacement ratio** of 60%, which is the pension benefit as a portion of final salary.

4.3.1 Pension Plan Portability and Job Mobility

DB plans are not portable. A plan is **portable** if benefits earned at one employer can continue to be accrued at another employer. In many cases, workers who die before retirement age receive no benefits from a DB plan and their heirs receive no lump sum or recurring benefit payments.

DB plans reward workers who spend their entire career with a single employer and punish workers who exercise job mobility. Contrast an employee who worked for 40 years at one firm to another employee who worked 20 years at each of two employers. Each employer provides a benefit of 1.5% of the average of the final five years of salary multiplied by the number of years of service. The worker started with an income of \$15,787 in 1975 and retired in 2015 with an income of \$50,000 after receiving annual salary increases of 3% over 40 years. If the worker served her entire career with one employer, the annual benefit would be \$28,302 (1.5% × 40 years × the final five-year salary average of \$47,171). The benefits would be quite different had she worked for two employers. The retiree worked at the first employer from 1975 to 1995, with an average annual salary in the final five years of \$26,117. The annual benefits of \$7,835 (1.5% × 20 years × \$26,117) are determined in 1995, but not paid until retirement in 2015. The second employer pays annual benefits in the amount of \$14,151 (1.5% × 20 years × \$47,171).

Compared to the annual benefit of \$28,302 after working the entire career for a single employer, the employee splitting careers between two firms earns an annual pension of only \$21,986 (\$7,835 plus \$14,151), which is \$6,316 per year less than if she had worked for a single firm.

A lack of portability may be an even greater issue for an employee who works a large number of jobs in a career, as many companies have vesting periods of five to 10 years. An employee must work for the entire vesting period to earn any retirement benefits. In a worst-case scenario, consider an employee who worked for 45 years, serving nine years at each of five employers. If each employer required a minimum of 10 years of service to qualify for a DB pension, the employee would have earned no retirement benefits, even after working for 45 years at firms offering DB plans.

4.3.2 Defining Liabilities: Accumulated Benefit Obligation and Projected Benefit Obligation

It can be challenging to model the liability of an employer's DB plan. Defining the liability is important, as employers need to reserve assets each year to plan for future benefit payments. A number of assumptions need to be made to calculate the amount owed in retiree benefits. These assumptions include:

- The amount of employee turnover and the years of service at the date of separation
- Average wages at retirement, which requires the current wage, estimated retirement age, and annual wage inflation from today until retirement
- The assumed age of worker death, as the number of years of benefits to be paid is the difference between the age at retirement and the age at death
- The number of current employees, hiring plans, and the anticipated age of all employees

The **accumulated benefit obligation** (**ABO**) is the present value of the amount of benefits currently accumulated by workers and retirees. This number may be very small for a young firm with young workers, such as a four-year-old technology start-up filled with young college graduates. In this scenario, current workers have had only four years to accrue benefits and the firm may not anticipate retirements for another 40 years. The ABO is relatively easy to calculate, as the number of workers, their tenure, and average salary are all known. Of course, future wage growth and the average employee life span need to be assumed.

The **projected benefit obligation** (**PBO**) is the present value of the amount of benefits assumed to be paid to all future retirees of the firm. This number is much more challenging to calculate, as the number of workers at the firm in the future, employee turnover levels, and years of service are unknowns. As long as the company has current employees, the PBO is always greater than or equal to the ABO. When the firm and its employees are young, the ABO may be much smaller than the PBO. For example, the PBO may assume 40 years of service whereas employees at the young firm have accrued only four years of service. In a mature firm with a large number of retirees and an older workforce, the ABO will be of a similar magnitude to the PBO. The difference between the ABO and the PBO is primarily based on the current versus future salaries and years of service of current employees.

4.3.3 Funded Status and Surplus Risk

The **funded status** of a pension plan is the amount of the plan's current assets compared to its projected benefit obligation (PBO) or accumulated benefit obligation (ABO). The funded status may be expressed in terms of currency, such as €2 billion underfunded, or in percentage terms, such as 70% funded (or 30% underfunded) if a plan's assets are 70% of the PBO. Plans should strive to be close to 100% funded. Overfunded plans, such as those with assets of 120% of PBO, may attract attention from employees who would like to earn larger benefits, or from corporate merger partners who may wish to disband the pension and keep the surplus value. Underfunded plans, such as those where assets are 70% of the PBO, may require larger employer contributions and may attract regulatory scrutiny.

The funded status of pension plans can vary sharply over time, as shown in Exhibit 4.2. The assets of the plan grow with employer contributions, decline with retiree benefit payments, and change daily with returns to the investment portfolio. The PBO also changes over time, as the present value factor is based on corporate bond yields. As corporate bond yields rise, the PBO declines. Conversely, declines in corporate bond yields lead to an increasing PBO.

The Citigroup Pension Liability Index tracks corporate bond yields that can be used to discount future values of the PBO. At December 31, 2009, the discount rate

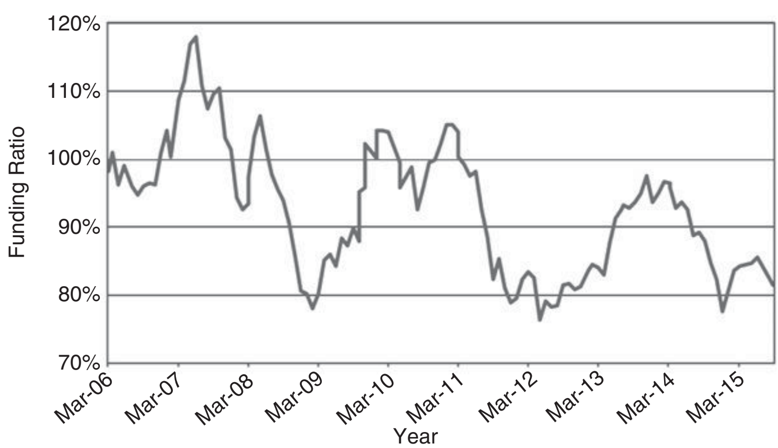


EXHIBIT 4.2 Estimated Funding Ratio of UK Pension Schemes
Source: *The Purple Book* (2015).

was 5.98%, while the duration of PBO benefits was estimated at 16.2 years. By year-end 2015, the discount rate had fallen to 4.34%. The pension plan's PBO can be compared to a short position in corporate bonds, which will change in value by the approximate amount of:

$$\%\text{Change in Liabilities} = -\text{Modified Duration} \times \text{Change in Yield} \qquad (4.2)$$

Over this six-year period, the 1.64% decline in corporate bond yields has led to an increase of 26.6% in the PBO, assuming that duration and future benefits assumptions remain unchanged.

$$26.6\% = -16.2 \times (-1.64)$$

The **pension surplus** is the amount of assets in excess of a pension plan's projected pension benefit (PBO). The **surplus risk** of a pension plan is the economic exposure to the spread between the assets and liabilities of a pension plan and can be measured as the volatility and tracking error of the difference between the value of the assets relative to the present value of the liabilities. Consider the example in Exhibit 4.3, where assets are invested 60% in the S&P 500 and 40% in the Barclays Aggregate Bond Index. The liabilities are assumed to have a duration of 16.2 years and a discount rate tracked by the Citigroup Pension Liability Index. From 1997 to 2015, the volatility of the asset portfolio was 11.0%, while the volatility of liabilities based only on the change in corporate bond yields was 10.4%. Because assets and liabilities had a correlation of −0.25 over this period, the surplus risk was even higher, as the volatility of the annual difference between asset and liability returns was 16.9%.

In summary, during times of low investment returns and rising amounts of regulatory pressure, many employers no longer choose to offer DB plans. Employers that provide DB plans will experience surplus risk on their balance sheets, and the

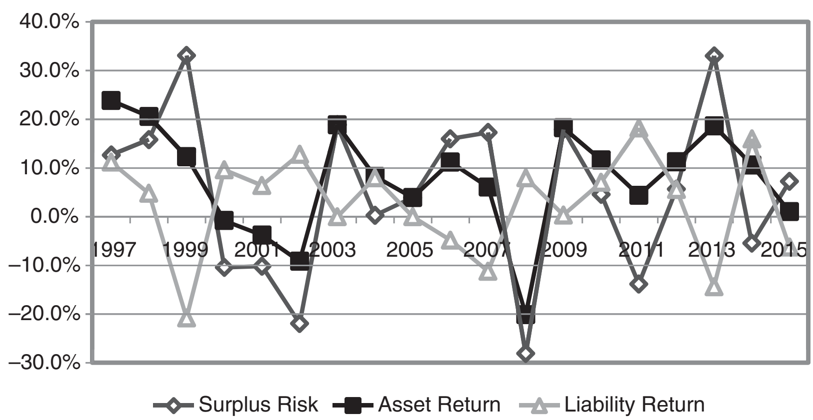


EXHIBIT 4.3 The Volatility of Pension Assets and Liabilities Creates Surplus Risk
Source: Authors' calculations based on returns to the S&P 500, Barclays Aggregate Bond Index, and the Citigroup Pension Liability Index.

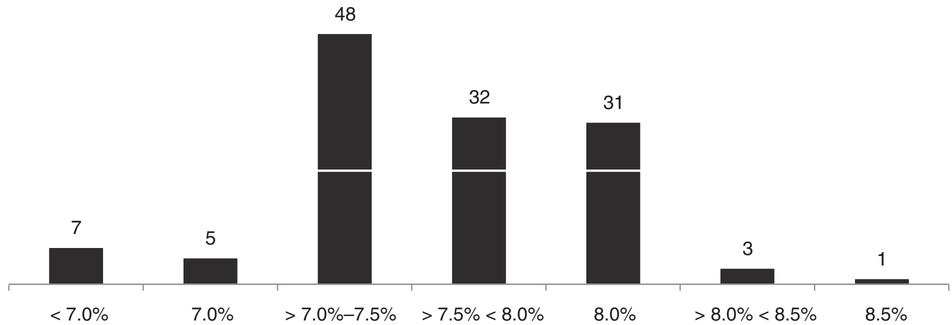


EXHIBIT 4.4 Distribution of Investment Return Assumptions, Fiscal Year (FY) 2015
Source: Public Fund Survey of the NASRA and the NCTR.

contributions to employee retirement plans will be variable. These employers will see pension costs rise during times of low investment returns; however, they will also earn the upside during times of high investment returns.

4.3.4 Why Defined Benefit Plans Are Withering

Each pension plan has a required return assumption that is used to calculate the employer's annual contribution. As shown in Exhibit 4.4, all of the 127 U.S. public pension plans surveyed by the National Association of State Retirement Administrators (NASRA) and the National Council on Teacher Retirement (NCTR) used return assumptions of 8.5% or less in 2015, with over 24% using an estimate of 8%. Note that return estimates have been falling in recent years, as 44% of pensions used an 8% return assumption in 2010. Should long-term investment returns fall below this assumed return, either the plan will become underfunded or additional employer contributions will be required. The required return is also a key driver of asset allocation, as investment policy is set in an attempt to earn the required return. That is, plans with higher required return assumptions may pursue a more aggressive asset allocation to earn the investment profits needed to justify both the current level of benefits promised as well as the employer contributions.

Plan sponsors, whether in the public or private sector, are increasingly becoming concerned about the affordability of DB plans. While corporate plan sponsors use a corporate bond yield as the discount rate, public plans use the required return assumption as the discount rate. The calculations underlying Exhibit 4.3 show an average annual return on assets of 7.2% from 1997 to 2015 while liabilities were increasing. When the public plan sponsor is making contributions based on an 8% required return and actual returns are substantially lower, funded ratios will decrease over time.

Regulatory changes, at least in the United States, are also making corporate DB plans less attractive. The Pension Protection Act of 2006 requires that corporate employers disclose the plan's funded status to plan participants. The Act also requires employer contributions to be commensurate with the funding status, with underfunded plans requiring greater contributions and overfunded plans requiring lower contributions. Underfunded plans must increase required contributions by an amount that projects the plan to be fully funded within seven years.

Investors are also concerned about the risk of investing in the equity securities of companies with underfunded pensions. The funded status of U.S. pension plans is now required to be disclosed on corporate balance sheets. Merton (2006) states that companies with large pension deficits may trade at lower multiples of earnings and book value, may exhibit higher betas, and may experience higher stock price volatility. The higher beta caused by pension risks can increase the firm's weighted average cost of capital by up to 2.7%, which makes it more difficult to find profitable operating investments.

Employees are also concerned about DB plans. The declining number of DB plans offered by companies and their lack of portability make such plans less relevant today. Employees are working at a greater number of firms during their careers than did previous generations. The U.S. Bureau of Labor Statistics (BLS) estimated in 2010 that Americans born between 1957 and 1964 held an average of 11 jobs before the age of 44, and only 12% of this population held four or fewer jobs during the first half of their career. It is therefore difficult for the majority of younger workers to accrue meaningful retirement income under a DB system.

As a result, as shown in Exhibit 4.5, DB plans are declining as a share of assets among U.S. pension plans.

Should a plan sponsor no longer wish to offer a DB plan to its employees, it has the option to freeze or terminate the plan. As a less drastic measure, the employer may move to a two-tier structure, offering newly hired employees a less generous pension benefit than previously hired employees. A **frozen pension plan** is one where employees scheduled to receive defined benefit (DB) pension benefits will no longer continue to accrue additional years of service in the plan. An employee with 20 years of service when the plan is frozen might retire five years later with 25 years of service, but the benefits would be tied to only 20 years of service. A **terminated pension plan** is no longer operated by the employer. Once a plan has been terminated, all assets will leave the control of the employer and either be paid out in lump sums to employees or be used to purchase annuities that will pay future benefits to retirees. Freezing or terminating pension plans is extremely popular in the United Kingdom, where *The Purple Book* (2014) estimates that only 13% of UK plans are open to new participants and allow current participants to continue to accrue benefits. Olsen (2012) states that "46% of U.S. corporate DB plans are active and open to new hires while 24% are closed, 24% are frozen, and 1% are being terminated."

4.3.5 Asset Allocation and Liability-Driven Investing

Pension plan sponsors have conflicting goals when designing the asset allocation of the plan. The first goal is to earn a high return on pension assets, which will be used to reduce the employer's long-term contributions required to fund employee benefits. The second goal is to minimize the degree of underfunding or the amount of surplus risk incurred in the plan.

As can be seen in Exhibit 4.6, Preqin (2016a, 2016b, 2016c, 2016d) surveyed a large number of institutional investors, noting the size of their allocation to alternative investments. Endowments and foundations (averaged here from Preqin data) are the leaders in allocating to alternative investments, but pension plans are not far behind. While endowments and foundations may have reached the maximum

EXHIBIT 4.5 Defined Benefit Assets of the Top 1,000 U.S. Pension Plans Are Losing Share over Time

	2002	2004	2006	2008	2010	2012	2014	2015
Total Plan Assets ($ Million)	$4,329,015	$5,351,019	$6,487,729	$6,395,807	$6,561,618	$7,534,167	$9,054,534	$8,843,028
Total DB Assets	$3,243,189	$3,969,566	$4,776,551	$4,618,163	$4,651,389	$5,242,362	$6,032,511	$5,828,983
Total Non-DB Assets	$1,085,826	$1,381,453	$1,711,178	$1,777,644	$1,910,229	$2,291,805	$3,022,023	$3,014,045
% Non-DB Assets	25.1%	25.8%	26.4%	27.8%	29.1%	30.4%	33.4%	34.1%

Source: Author's calculations, *Pensions & Investments*.

EXHIBIT 4.6 Average Assets Allocated to Alternative Investments by Institutional Investors

	Private Pensions	Public Pensions	Sovereign Wealth Funds	Endowments and Foundations
Private Equity	5.4%	5.8%	6.3%	11.0%
Hedge Funds	11.0%	8.3%	6.2%	19.0%
Real Estate	7.9%	9.0%	7.4%	6.0%
Infrastructure	2.5%	3.2%	NA	3.5%
Total	26.8%	26.3%	19.9%	39.5%

Source: Preqin (2016).

practical allocation to alternative investments, it appears that pension plans are continuing to grow their allocations to alternative investments. Jacobius (2015) reports that the largest 200 defined benefit pension plans had invested nearly $900 billion in alternative investments by 2014, double the level in 2009.

While investing in equity and alternative investments may earn higher long-term returns, these risky assets are subject to substantial short-term volatility, whether measured against a benchmark of zero, the plan's required return, or the change in the present value of the plan's liabilities. Companies wishing to reduce surplus risk may have a very large fixed-income allocation. While this reduces surplus risk, the large fixed-income allocation reduces the likely return on assets, which increases the plan sponsor's long-term contributions, assuming the plan is less than fully funded.

Liability-driven investing (LDI) seeks to reduce surplus volatility by building a portfolio of assets that produces returns that are highly correlated with the change in the plan's liabilities. The simplest way to immunize pension liabilities is to invest in a corporate bond portfolio with a duration matching that of the liabilities. Other ways to reduce surplus risk include derivatives overlays, such as a swap receiving long-duration bond returns or a swaption that increases in value as interest rates decline. (See Exhibit 4.7.)

Meder and Staub (2007) discuss the asset allocation necessary to hedge the ABO and PBO exposures. The ABO does not count future benefit accruals; it simply has exposure to declining nominal bond rates, which increases the present value of benefit payments. The PBO is more complicated, as future wage inflation may be correlated to both equities and inflation rates. Employers with younger workers would have a higher allocation to equities. Plans may offer retirees a **cost of living adjustment (COLA)**, which increases the benefits paid to employees along with the rate of inflation. For example, consider a retiree earning a pension of $2,000 per month. After five years of 4% inflation rates and a 75% COLA (75% of 4%), the retiree's pension will have risen by 3% per year to $2,318 per month. Plans offering benefits with large percentage COLA adjustments would need to have large allocations to inflation-protected bonds, in order to reduce surplus risk. **Inflation-protected bonds** earn a nominal coupon while the principal value rises with the rate of inflation. Due to the superior hedging capabilities of real bonds, their total return tends to be very low.

Investors, then, may wish to protect their portfolios against inflation without earning the low real returns offered by inflation-protected bonds. A growing number of investors are turning to real assets to gain inflation protection while

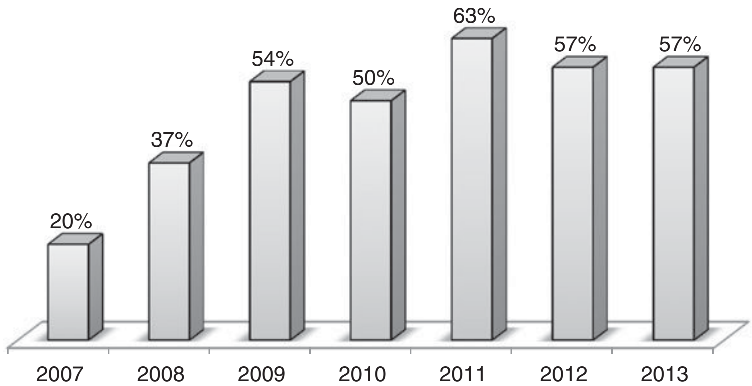


EXHIBIT 4.7 Poll Results of Pension Management Research Panel Liability-Driven Investing Survey
Source: "7th Annual Liability-Driven Investing (LDI) Poll: More Plan Sponsors Using LDI Than in Years Past," SEI Institutional Solutions, December 2013.

attempting to earn higher returns than offered by inflation-protected bonds, such as Treasury Inflation-Protected Securities (TIPS). In addition to the automatic inflation protection offered by TIPS, Martin (2010) demonstrates that a number of real assets can serve as long-horizon inflation hedges, including commodities, timber, and farmland. While equities are not a good hedge against long-horizon inflation, it is uncertain whether infrastructure, real estate, or intellectual property investments are good hedges against inflation. Specific investment characteristics, such as fixed-rate debt and revenues tied to inflation, improve the ability of infrastructure or real estate to serve as an inflation hedge. Assets with fixed-rate leases and variable-rate debt may actually be hurt by inflation, even though they are real assets, such as real estate or infrastructure.

4.4 GOVERNMENTAL SOCIAL SECURITY PLANS

Government social security plans may provide retirement income to all previously employed citizens of a given country, regardless of whether the worker was employed in the public sector or in the private sector. The main requirement for earning benefits from these systems is that retirees must have worked for a minimum amount of time, such as 10 years over the course of a career, and paid contributions into the system. Social security benefits are typically portable, meaning that employees continue to accrue service credits whenever they are paying contributions into the system, regardless of the number of employers in a career. Some employees, especially of governmental entities, do not receive these benefits, as neither employees nor their employers paid the required contribution. Whitehouse (2007) estimates that the average retirement benefit in 24 high-income Organization for Economic Cooperation and Development (OECD) countries is 31% of average earnings.

DB plans often have benefits explicitly tied to employee income, without a cap on the amount of benefits that may be earned. Social security plans are quite different, in that there are caps on earnings, which means that retirees with lower career-average incomes may earn a higher retirement income-replacement ratio than higher-income retirees. This is a **progressive system**, where lower paid workers get relatively higher benefits than those earned by higher paid workers when measured by a percentage of salary. U.S. workers retiring in 2016 at the age of 66 were eligible for a maximum monthly retirement benefit of $2,639. This maximum benefit is paid to higher-income workers, such as those with incomes over $118,500 in 2015. According to a 2015 study by the U.S. Congressional Budget Office (CBO), workers born in the 1960s are estimated to have a retirement income-replacement ratio of 82% for the lowest quintile of U.S. workers, which declines to just 23% for the highest quintile of workers based on their five highest years of earnings. Social security systems may also provide income security to the dependents of workers, paying benefits to the spouse or children of workers who die or become disabled during their working years.

4.5 CONTRASTING DEFINED BENEFIT AND CONTRIBUTION PLANS

Willis Towers Watson (2016) estimates that global pension assets in 2016 were 51.6% defined benefit (DB) plans and 48.4% defined contribution (DC) plans. The mix between DB and DC plans varies widely by country, with 87% of Australian assets invested in DC plans, whereas 96% of Japanese assets are in DB plans. Using Investment Company Institute (ICI) (2015) estimates, assets in U.S. DB and other plan assets of $10,400 billion are now smaller than the combined $6,800 billion in DC assets and $7,400 billion in individual retirement accounts.

4.5.1 Defined Contribution Plans

The most common alternative retirement plan offered by employers is a **defined contribution plan**, where the employer makes a specified contribution on behalf of each covered employee on a regular basis such as a percentage of the employee's salary. In a DC plan, there is no surplus risk for the employer, as assets always match liabilities. A common structure for a DC plan is one in which an employer offers each employee an annual amount of 3% of salary, with perhaps a matching contribution of 50% of the amount contributed by the employee. A **matching contribution** is a voluntary contribution made by an employee that is augmented by additional contributions by the employer. For example, an employee will contribute 6% of salary to the DC plan while the employer will contribute 3% plus a matching contribution of 3% (50% of 6%). This employee would place 6% of salary into a retirement account, which, combined with the firm's contribution of 6%, totals a 12% contribution.

4.5.2 Plan Differences in Portability, Longevity Risk, and Investment Options

In contrast to DB plans, DC plans are portable, meaning that the employer contributions become the asset of the employee once the vesting period is completed.

This portability is better for employees who work multiple jobs in a career, and for employees who work for firms that may not have the financial strength to pay long-term pension benefits in the amount promised. When leaving an employer, the employee is able to roll over the balance in the DC plan into the plan offered by the next employer or into an individual retirement account. Given that DC plans are personal accounts, the employee contribution, investment gains, and vested portion of employer contributions can be given to the employee's heirs should the employee die before retirement.

In a DB plan, the longevity risk is incurred by the employer. The employee is guaranteed the monthly benefit for life, whether that life is longer or shorter than anticipated. This means that employees cannot outlive their assets. In a DC plan, however, there is no guarantee as to the amount of assets accumulated or the amount of monthly income in retirement, meaning that longevity risks directly impact the employee. Employees with low contributions, low investment returns, or long lives may have a significant probability of "living too long," meaning that their assets may be exhausted or their spending rate curtailed in their final years of life. Employees need to plan for at least 20 years of retirement income, as Maginn et al. (2007) estimate that, in the United States, a 65-year-old couple has a 78% chance that at least one of them will live beyond age 85.

The employer makes the asset allocation decisions in a DB plan, but asset allocation decisions in DC plans are made by the employees, typically using the fund choices provided by the employer. The employer may offer a range of investment choices, such as up to 20 mutual funds. Kupperman and Kilgallen (2015) report that the average defined benefit plan outperformed the average defined contribution plan by an annual average of 140 basis points from 1997 to 2011. Waring and Siegel (2007) note that DC plans have higher fees, with the average fund in a DC plan charging over 100 basis points whereas fees in DB plans average less than 50 basis points per year.

4.5.3 Asset Allocation in Defined Contribution Plans

It is the employee's decision as to how much to contribute to the retirement account, as well as how to allocate the assets across the allowed investment choices. Leaving the decision-making to employees, most of whom are not trained in making investment decisions, can lead to a wide variety of employee outcomes. Some employees may retire without any retirement assets, either because they did not choose to participate in the plan or because they were allowed to invest all of their assets in their employer's stock, which ended up worthless at the end of a bankruptcy proceeding. Some employers may offer the option of a brokerage window, which allows employees to invest in a broader variety of mutual funds, or even individual stocks. While employees with a high degree of financial sophistication can benefit from a brokerage window, the sheer number of options or the ability to concentrate risk in more narrow investments can cause excessive risk for some plan participants.

On the other hand, a diligent saver with good investment returns can potentially earn a larger retirement benefit in a DC plan than in a DB plan. In the earlier example, the employee started with a salary of $15,787 and worked for 40 years, before retiring at a final salary of $50,000 with a single-employer DB pension plan income of $28,302. The same employee, when covered by a DC plan, could have invested 6% of her salary and earned employer contributions in a similar amount. Assuming salary raises of 3% per year, investment returns of 8% per year, and annual contributions in the amount of 12% of salary, the employee would have accumulated more than $699,000 at retirement. This amount includes employee and employer contributions of approximately $71,400 each over the course of the career and over $550,000 in investment earnings. With a spending rate of just 4.05%, the employee would earn the same amount as the DB pension plan income of $28,302. If the DC account earned annual returns of at least 4.05% during retirement, the nominal value of the retirement account would either be stable or rising for the rest of the employee's life. The principal balance, at the date of death, would be passed on to the retiree's heirs.

In contrast, most DB plans do not offer value to the family of the retiree, unless there is a promise to pay some portion of the pension income for the rest of the spouse's life. For a diligent saver who was blessed with high investment returns over the course of a career, the DC plan is far superior to a DB plan, in terms of both portability and the ability to pass significant assets along to heirs. Unfortunately, many DC plan participants either save too little or invest too conservatively, and end up faced with the prospect of earning far fewer benefits from the DC plan than if the employee had worked for a single employer offering a DB plan. When plan participants have the ability to withdraw from or borrow against the assets in the DC plan before retirement, it becomes even more difficult to accumulate the assets necessary to ensure a strong income during retirement.

Given that employees are making their own investment decisions, many employers offer simple fund choices for DC plan participants. In some cases, employees are allowed only the choice to allocate assets across domestic stocks, domestic bonds, cash, and global stocks. Other plans will allow participants to invest in equity securities of the employer, as well as more than 20 funds in a variety of geographies or asset classes.

Employees do not generally allocate DC plan assets in the same careful way that professional managers allocate DB plan assets. Employees often invest in just a single fund, resulting in a 100% equity or 100% cash allocation, or they diversify contributions equally across all investment choices. Employees also do not frequently rebalance or change allocations when their investment needs become more conservative as retirement approaches. This lack of rebalancing results in a **drifting asset allocation**, where the allocations to asset classes change based on returns of each asset class with the highest-returning asset classes growing as a share of the portfolio. For example, an employee may have decided at age 30 to direct 70% of contributions to an equity fund and 30% to a fixed-income fund, given his investment needs at the time. If stock returns were substantially higher than bond returns over the next 20 years, the now 50-year-old employee may find himself with a portfolio with 85% equity and 15% fixed-income allocations at a time when a 60% equity, 40% bond mix may be more appropriate for his circumstances.

4.5.4 Target-Date Funds and Alternative Investments within Pension Plans

Given that most DC plan participants are not accredited investors, few DC plans offer private placement alternative vehicles such as private equity. The limited number of fund choices in DC plans reduces diversification potential and the ability to switch to funds with better performance. Due to a lack of investor sophistication as well as regulatory restrictions, most DC plan participants do not have the ability to invest directly in alternative investments. When alternative investment choices are offered in DC plans, they typically focus on commodities or real estate. It is quite rare for individual employees to be able to invest in private equity or hedge funds through DC plans.

In the United States, after the Pension Protection Act of 2006, many employers made changes to their DC plans' designs in order to alleviate a number of problems with DC plan investments. In the past, employers may not have mandated DC plan participation, and when they did, all employee contributions were placed in cash, unless otherwise directed by the employee. Recently, employers have been automatically enrolling new employees in DC plans, setting the employee contribution at 1% to 3% of salary, and automatically increasing annual contributions by one-third of the employee's salary increase. For example, an employee contributing 2% of salary would have a contribution rate of 4% after earning 3% salary increases for two years. Finally, the default investment option may now be a target-date fund, rather than cash.

A **target-date fund** has risks that are managed relative to a specified horizon date, which allows employees to choose a single investment option without worrying about rebalancing or changing investments as the horizon date approaches. A young employee hired in 2015 might invest in a target-date fund, anticipating retirement in the year 2050, while an employee approaching retirement may be invested in a fund targeting a retirement date of 2020. The 2050 target-date fund assumes that a young employee with an average risk tolerance may be invested 85% in equity and 15% in fixed income. This asset allocation would be regularly rebalanced by the fund manager, becoming more conservative over time. Ultimately, the 2050 fund (in 2045) would resemble the 2020 fund (today), with an asset allocation of 50% equity and 50% fixed income, matching the investment needs of an individual approaching retirement. Target-date funds are often managed as a fund of funds structure, with a mutual fund company allocating assets to between three and 20 mutual funds managed by the mutual fund firm. In this structure, funds with private equity, hedge funds, commodity, or real estate investments may be included in the target-date products at allocations between 5% and 20%. The ICI (2015) estimates that over 10% of DC plan assets in 2015 were invested in target-date funds.

While most DC plans do not offer alternatives as investment options, Kupperman and Kilgallen (2015) suggest that target-date funds investing in liquid alternative investments are likely to grow the ability for DC investors to allocate to alternatives. However, alternative investments are more likely to be held as an allocation within the professionally managed target-date funds than to be made available to plan participants as stand-alone investment options.

4.6 ANNUITIES FOR RETIREMENT INCOME

This section discusses annuities, which have two related major definitions. An annuity is often defined as any fixed stream of cash flows. Annuities also refer to specific insurance products that contain features that enable payouts in the form of a stream of cash flows.

4.6.1 Financial Phases Relative to Retirement

When individuals are working, they are in the **accumulation phase** of their lives where they are saving a portion of their income and growing their assets to provide for a comfortable retirement. After retirement, the ability to earn income declines substantially, and the investor enters the **decumulation phase**, where assets are drawn down to support spending during retirement. In the accumulation phase, income should exceed spending, which allows the retiree to have spending greater than income. Assets decumulated during retirement can draw from the worker's own savings or from a pension or social security program.

Each retiree will have a desired level of income, which can be drawn from any of these sources. Some retirees who saved and invested well during their careers or are the beneficiaries of generous pension plans may find that the resources available to them are unlikely to be exhausted in their lifetimes, given a reasonable spending rate. These retirees will not outlive their assets, and will be able to bequeath assets to their heirs or charitable causes.

Most retirees, however, will not have the luxury of dying with a significant amount of assets in savings and investments. Some investors fear longevity risk, in that a longer than expected lifetime will be highly likely to deplete their assets when spending at the desired rate.

4.6.2 Three Important Risks to Retirees

Starting with a lump sum of assets at retirement, investors face at least three important risks. First is longevity risk, where the longer the life of the retiree, the more likely he is to deplete his assets. Second, there is the market risk of the investments, where a substantial investment loss will deplete the retiree's assets faster than the desired rate. Third, there is the risk of inflation, where a fixed amount of spending each year buys fewer goods and services in a rising inflation rate environment.

Actuaries calculate **mortality tables**, which show the distribution of the expected age of death or probability of death for various current ages across a specified population. With advances in medicine, people have been living longer. For example, Coughlan (2014) reports that a 65-year-old male living in the United States had an expected remaining life expectancy of 14.2 years in 1951 and 18.7 years in 2014. Mortality tables can be quite accurate for large populations of people, but can err substantially as the sample size gets smaller or the health or occupation of the measured lives changes. The maximum error in mortality tables is likely to be on the life of a single person, which can be much longer or shorter than expected, even if it is in the range of the population distribution of mortality.

All three of these risks are maximized when investors choose to manage their own assets in retirement. Ideally, an investor would be able to enjoy strong investment returns during his retirement, allowing him to have a sustainable spending rate and still leave assets to his heirs. Consider an investor with \$1,000,000 in assets at retirement and a need for \$40,000 in annual income to support his expenses, which are likely to grow at the rate of inflation over time. Should he earn 6% annual returns on his investment portfolio while spending 4% of assets each year and experiencing a 2% inflation rate, he can meet his annual spending rate after inflation and still leave the real value of the \$1,000,000 to his heirs. This same investor, however, would have a substantial problem if investment returns averaged zero over time with inflation of 4%. Should this investor (the U.S. male previously described) spend an inflation-adjusted \$40,000 each year, he will have exhausted his entire retirement savings at his life expectancy of 83 years old. Should he live longer than his expected lifetime, he will have experienced substantial longevity risk, living longer than his assets have actually lasted.

4.6.3 Estimating Exposure to Longevity Risk

A retiree could use the information provided by mortality tables along with some financial data about her retirement fund and the cost of living to obtain rough estimates of her exposure to longevity risk. The expected economic life of a retirement fund can be approximated by the following expression:

$$EL = -\frac{1}{\ln(1+R)} \times \ln\left(\frac{\text{Payment} - R \times \text{Assets}}{\text{Payment}}\right) \tag{4.3}$$

Here, EL is the expected economic life of the retirement fund measured in years, R is the annual net of fees real rate of return (i.e., after inflation) that the assets are expected to earn each year, Payment is the first annual withdrawal, which is expected to increase at the annual rate of inflation, and Assets is the current value of the fund's assets. Note that Equation 4.3 is valid as long as $R \times A <$ Payment. If the annual cash generate by the investment is greater than the annual payment, then the fund will last forever.

APPLICATION 4.6.3

Suppose that the market value of the retirement fund of an individual is €2,500,000. The annual net of fees real rate of return on the fund is estimated to be 3%, and the retiree wants to withdraw €200,000 in the first year, and an amount each year thereafter that will grow at the rate of inflation. What is the expected economic life of the fund?

Using Equation 4.3:

$$EL = -\frac{1}{\ln(1+0.03)} \times \ln\left(\frac{200{,}000 - 0.03 \times 2{,}500{,}000}{200{,}000}\right) = 15.9 \text{ Years}$$

The fund is expected to last 15.9 years.

Retirees in good health with relatively small amounts of retirement savings may have a good probability of exhausting their assets during their lifetimes. Those retirees may wish to purchase an immediate annuity or a deferred annuity from an insurance company. Sweeting (2014) reports that annuitization of retirement assets is compulsory in the United Kingdom, the Netherlands, Germany, and Italy. While voluntary annuitization may appeal to the most healthy retirees who hope to live longer than the mortality tables suggest, compulsory annuitization avoids this moral hazard and makes insurance company experience closer to the general population mortality tables.

4.6.4 Two Major Types of Annuities

In an **immediate annuity**, an investor pays a lump sum to an insurance company for cash flows starting in the first year of the contract and guaranteed for some period. Some retirees may purchase a lifetime income annuity, which pays a set amount of income for life. While the use of a lifetime income annuity reduces longevity risk by guaranteeing that the retiree will never run out of income, the entire cost of the annuity is paid to the insurance company with no residual value available to heirs. That is, whether the investor dies in one day or 30 years, no assets are left over after the investor has received income. This desire to leave assets for heirs has led to a low popularity of annuities in the United States.

Ezra (2016) advocates the use of a deferred annuity, which he calls longevity insurance. In a **deferred annuity**, an investor pays a lump sum to an insurance company for cash flows that are scheduled to start at some date in the future. Because the cash flows from the insurance company start at some future date, perhaps in 11 years, the cost of the deferred annuity is lower than the cost of the immediate annuity. Investors choosing a deferred annuity can keep some portion of their assets in an investment portfolio designed to provide income for 11 years or longer, while using the deferred annuity to provide lifetime income at an advanced age when longevity risk is experienced and when they are more likely to deplete their resources.

4.6.5 Analysis of the Value of a Growth Annuity

Consider the present value of an ordinary annuity, where *r* is the discount rate, and *n* is the number of annual payments. Several present values of an ordinary (immediate) annuity for various values of *r* and *n* are provided in the first column of Exhibit 4.8 based on Equation 4.4.

$$\mathrm{PV}_{\text{Ordinary Annuity}} = \frac{\text{Payment}}{r} \times \left[1 - \frac{1}{(1+r)^n}\right] \tag{4.4}$$

Now consider the present value of a growing annuity where *r* is the discount rate, *n* is the number of annual payments, and *g* is the annual growth rate of the payments as depicted in Equation 4.5. A growing annuity has payments that increase at the annual rate of *g*.

$$\mathrm{PV}_{\text{Growing Annuity}} = \frac{\text{Initial Payment}}{r-g} \times \left[1 - \left(\frac{1+g}{1+r}\right)^n\right] \tag{4.5}$$

EXHIBIT 4.8 Cost of Purchasing an Annuity to Provide $1,000 in Annual Income

	Immediate Fixed Annuity	11-Year Deferred Fixed Annuity	Immediate Annuity with Income Growing at 3% Inflation
20-Year Annuity, 5% Discount Rate	$12,462	$ 7,651	$15,965
20-Year Annuity, 2% Discount Rate	$16,351	$13,413	$21,546
30-Year Annuity, 5% Discount Rate	$15,372	$ 9,437	$21,920
30-Year Annuity, 2% Discount Rate	$22,396	$18,372	$34,002

APPLICATION 4.6.5

A pension plan offers a choice of a fixed $100,000 initial annual benefit or an $80,000 initial annual benefit that will grow by 3% per year. Using a market interest rate of 5%, calculate the present value of both alternatives based on life expectancy of 20 years (based on annual compounding).

Both alternatives can be valued using Equation 4.5 by using a zero growth rate for the fixed payment alternative and 3% for the growth alternative (as well as the different initial payments).

$$PV_{\text{Growing Annuity}} = \frac{\text{Initial Payment}}{r - g} \times \left[1 - \left(\frac{1+g}{1+r}\right)^{n}\right]$$

PV of Fixed = $2,000,000 * (1 – 0.376889) = $1,246,222
PV of Growing = $4,000,000 * (1 – 0.680704) = $1,277,184

The growth annuity has a slightly higher value. Note that the present value annuity formulas are highly intuitive and easily memorized. An annuity may be viewed as a perpetuity that is received today and lost after *n* years. The first term on the right side of the equation is the initial value of a perpetuity. The last term subtracts out the future value of the perpetuity. This interpretation can be seen in the solutions shown where the perpetuity values ($2 million and $4 million) are reduced for the payments not received after 20 years.

Comparing the cost of the immediate, deferred, and immediate growing annuities in Exhibit 4.8 for various discount rates and payment periods gives an important illustration of interest rate risk, inflation risk, and longevity risk for a retiree or a pension plan. It also illustrates the reduced cost of a deferred annuity relative to a growing annuity.

When designing a pension plan or a personal investment strategy, investors need to be aware of the cost of market risk, longevity risk, and inflation risk. Regarding interest rate risk or market risk, compare the cost of an immediate fixed-rate annuity with 20 annual payments using a 2% discount rate relative to a 5% discount rate. As interest rates decline from 5% to 2%, the cost of providing $1,000 in annual income rises from $12,462 to $16,351 per retiree. Longevity risk can also be illustrated, as the cost of providing $1,000 payments at a 5% discount rate for 20 years is $12,462 while the present value of 30 annual payments is $15,372. Finally, compare the cost of a growing annuity to a fixed annuity. For 20 annual payments with a 5% discount rate, adding 2% annual inflation benefits increases the cost from $12,462 to $15,965 per retiree.

The cost of a lifetime annuity would typically be based on the life expectancy and discount rate assumed by the insurance company, likely comparable to the 20-year example for a 65-year-old male shown earlier.

4.7 CONCLUSION

It is in everyone's best interest that members of our family, our colleagues, and our neighbors are able to have a retirement income sufficient to enjoy a comfortable lifestyle. That retirement income can be sponsored by a government benefits plan, a corporate defined benefit plan, or where an individual has more control over their investments through personal savings or a defined contribution plan.

When planning investments to fund retirement income needs, investors need to consider the market risk of their investments, the interest rate risk of their liabilities, inflation risk, and the expected lifetime of the retiree. Before promising benefits to retirees, plan sponsors should consider the structure of those benefits to ensure that they meet the needs of retirees as well as being affordable and sustainable for the plan sponsor.

NOTES

1. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTPENSIONS/0,,contentMDK:23231994~menuPK:8874064~pagePK:148956~piPK:216618~theSitePK:396253,00.html.
2. See "Pension Risk Management," Cambridge Associates LLC, Boston, MA, 2011.

REFERENCES

Congressional Budget Office. 2015. "CBO's 2015 Long-Term Projections for Social Security: Additional Information."

Coughlan, G. 2014. "First Steps in Longevity Risk Management." *Pension and Longevity Risk Transfer for Institutional Investors* (Fall).
Ezra, D. 2016. "Most People Need Longevity Insurance Rather than an Immediate Annuity." *Financial Analysts Journal* 72 (2): 23–29.
Investment Company Institute. 2015. *2015 Investment Company Fact Book*.
Jacobius, A. 2015. "Infrastructure, Energy Rise to the Top of Pension Fund Alts Portfolios." *Pensions & Investments*, February 9.
Kupperman, D., and S. Kilgallen. 2015. "The Case for Liquid Alternatives in Defined-Contribution Plans." *Journal of Alternative Investments* 18 (1): 59–66.
Maginn, J., D. Tuttle, D. McLeavey, and J. Pinto. 2007. *Managing Investment Portfolios: A Dynamic Process*. 3rd ed. Hoboken, NJ: John Wiley & Sons.
Martin, G. 2010. "The Long-Horizon Benefits of Traditional and New Real Assets in the Institutional Portfolio." *Journal of Alternative Investments* 13 (1): 6–29.
Meder, A., and R. Staub. 2007. "Linking Pension Liabilities to Assets." *Society of Actuaries*.
Merton, R. 2006. "Allocating Shareholder Capital to Pension Plans." *Journal of Applied Corporate Finance* 18 (1): 15–24.
Olsen, K. 2012. "46% of Corporate Defined Benefit Pension Plans Are Open." *Pensions & Investments*, January 10.
Pension Protection Fund. 2014. *The Purple Book*.
———. 2015. *The Purple Book*.
Preqin. 2016a. Global Hedge Fund Report.
———. 2016b. Global Infrastructure Report.
———. 2016c. Global Private Equity and Venture Capital Report.
———. 2016d. Global Real Estate Report.
Sweeting, P. 2014. "The Future of Pensions: A Plan for Defined Contribution." JPMorgan.
Waring, M. B., and L. B. Siegel. 2007. "Don't Kill the Golden Goose! Saving Pension Plans." *Financial Analysts Journal* 63 (1): 31–45.
Whitehouse, E. 2007. *Pensions Panorama: Retirement-Income Systems in 53 Countries*. Washington, DC: International Bank for Reconstruction and Development/World Bank.
Willis Towers Watson. 2016. "Global Pension Assets Study."

CHAPTER 5

Sovereign Wealth Funds

This chapter covers Sovereign Wealth Funds (SWFs), which are rapidly amassing assets to become one of the world's largest categories of investors in traditional as well as alternative assets. The CAIA Level I textbook defines sovereign wealth funds as state-owned investment funds held for the purpose of future generations and/or to stabilize the state currency. SWFs controlled over \$7 trillion at the end of 2015, which is more assets than the combined AUM of the entire hedge fund and private equity industries. With over \$1.4 trillion invested in alternative assets, it is important to understand the investment goals of SWF investors.

5.1 SOURCES OF SOVEREIGN WEALTH

The creation of a SWF generally results from a nation having an excess of savings, usually a result of substantial budget surpluses over the course of many years, whether from favorable macroeconomic conditions, planning, and/or restraint. More specifically, the source of such wealth is often associated with natural resources (usually oil and gas, and, to a lesser extent, copper, diamonds, and phosphates), strong and reliable trade surpluses, or the receipt of foreign aid money. In fact, two broad categories are responsible for nearly all SWF assets: oil and/or gas (\$4.2 trillion) and non-commodity (\$3.0 trillion) sources.

5.1.1 Accounting for Changes in the Reserve Account

The **reserve account** of a central bank consists of the central bank's holdings of foreign currencies and is operated by the central bank to conduct transactions involving foreign currencies. A country's balance of payments measures the inflows and outflows of currency that facilitate each country's foreign transactions. The balance of payments considers three accounts, which must offset each other in any given year as depicted in Equation 5.1:

$$\Delta\text{Reserve Account} = \Delta\text{Current Account} + \Delta\text{Capital Acccount} \tag{5.1}$$

Where ΔReserve Account is the change in the reserve account, and Current Account and Capital Account are positive numbers for surpluses and negative numbers for deficits. For example, a country with a €300,000 current account surplus and a €200,000 capital account deficit would experience a €100,000 increase in its reserve account.

APPLICATION 5.1.1

Country A experienced a €100,000 drop in its reserve account while having a €300,000 surplus in its capital account. What was the current account surplus or deficit of country A? Inserting the known values into Equation 5.1 and solving for the current account surplus generates a value of –€400,000, which indicates a current account deficit of €400,000.

The current account measures trade in goods and services along with investment income and gifts, such as foreign aid. A country is said to have a **current account deficit** when the value of its imports of goods and services exceeds the value of its exports, meaning that more currency is flowing out of the country to purchase these goods and services than is flowing in from selling goods and services. The capital account measures investment flows, where foreigners buy assets in a country or make loans. A **capital account surplus** occurs in a country when the amount of imported capital exceeds the amount of exported capital.

If a country regularly imports more goods and services than it exports, it runs a current account deficit and is dependent on capital account transactions continuing to flow into the country to finance those deficits. For example, the United States tends to run a current account deficit and a capital account surplus. Residents of the United States tend to buy a far greater amount of imported goods and services than the U.S. economy exports. In order to finance this excess consumption, U.S. residents need to sell assets or borrow money. China has the opposite profile, with a current account surplus and a capital account deficit, as they export goods and use the proceeds to increase their ownership of foreign assets.

If there is a situation of capital flight and a capital account surplus turns into a capital account deficit while the country still has a current account deficit, the country's official reserve account may need to be quickly drawn down to maintain the balance of payments. This was the cause of the 1997 currency crisis that led to large declines in the currencies of Thailand, Malaysia, and Indonesia as described in Chapter 17 of the CAIA level I textbook. Some attribute the rise of SWFs as a response to this crisis, where countries with fixed rate currencies had insufficient reserves to prevent a crisis.

5.1.2 Changes in the Reserve Account and Five Drivers of Currency Exchange Rates

In the case of a free floating currency, a capital account surplus tends to roughly offset a current account deficit in any given year, as the currency price will adjust to make imports of goods or exports of capital equally attractive. If a country's exports seem to be cheaply priced, the amount of exports is likely to increase, which creates demand for that country's currency that is likely to lead to the currency's appreciation. When the currency is freely floating, the change in the government's foreign currency reserves tends to be small, as the currency market tends to bring equilibration to the capital account and the current account.

Like any other asset, currency prices are typically set by supply and demand. Gwartney et al. (2003) explain that a country's currency is likely to appreciate when the country has:

1. A lower inflation rate than its trading partners
2. Higher real interest rates than its trading partners
3. Policies that attract an inflow of capital
4. Slower income growth versus trading partners that reduces the demand for imports
5. A competitive or comparative advantage in export-oriented industries

These five values tend to drive changes in the value of the currency of a country with a free floating currency. For example, a country's free floating currency is likely to appreciate in value when foreigners buy the assets and exports of that country at a greater rate than that country's residents buy the assets and exports of other countries.

Central banks can attempt to manage the value of its currency or fix the value of its currency relative to another currency. For example, suppose that a central bank is able to lower the value of its currency by 10% relative to other currencies while the nominal prices of goods and services are constant. The prices of imports to that country will be roughly 10% higher in terms of the domestic currency to that country's citizens while exports from that country will be roughly 10% cheaper to citizens of other countries when measured in their currencies. The intended effects are to expand the country's exports and cut the country's imports.

When currency prices are not allowed to adjust freely due to market forces, the official reserve account can grow or shrink, as the current account and the capital account will not naturally balance. For example, the value of the Chinese yuan was pegged at 8.28 to the U.S. dollar from 1994 until 2005. Over much of this period, it appeared that Chinese labor costs were low and the yuan was undervalued. With low labor costs and an undervalued currency, China built a very large export economy, which created demand for the yuan. As foreign residents purchased these Chinese exports, they created a demand for the undervalued yuan. In order to support the currency at the pegged rate, the Chinese government was selling yuan at its undervalued price and buying foreign currency, predominately U.S. dollars and euros. After starting a managed float in 2005, the yuan had appreciated to 6.20 to the U.S. dollar by 2015 and the Chinese official reserve account had swelled to over $4.1 trillion. Some of these foreign currency reserves can be held by the central bank to enact interest rate and currency stabilization policies, while a portion of the reserves can be transferred into a SWF for long-term investment purposes. By keeping the yuan undervalued, Chinese exports were underpriced and imports overpriced, creating greater incentives for Chinese consumers and businesses to save. The Chinese government then channeled these savings into their SWFs.

Sovereign wealth funds of countries with persistent current account surpluses, such as China, Hong Kong, and Singapore, now control over $2.6 trillion. One reason why countries may wish to amass foreign currency reserves is that flows in the capital account and the current account tend to move at different speeds. Foreign trade flows and foreign direct investment can move rather slowly, as the value of the currency takes time to impact trading and real asset investment decisions. However, portfolio investment flows can move much more quickly. This makes it dangerous for a country to run a current account deficit and a capital account surplus.

5.1.3 Commodity Exports and the Reserve Account

Current account flows can be differentiated into flows from manufactured goods, services, and commodities. While many countries, especially in Asia, export goods and services to earn foreign currency revenue, other countries, especially in the Middle East, earn foreign currency revenues through the export of commodities, such as oil, gas, and copper. Economies that export goods and services may feel comfortable at projecting similar or growing export revenues from year to year, as it can be relatively easy to project prices and quantities of manufactured goods. Commodity producers, however, have no such certainty. Crude oil prices have a higher volatility than that of most equity markets and many other commodities, which can make it difficult for a country to predict the value of their exports in any given year. The price of crude oil averaged less than $11 per barrel in 1998, rose to over $90 per barrel in 2008, and fell below $40 per barrel by the end of 2015.

Government revenues in many countries are tightly linked to oil prices, either because the oil company is state owned or because the state earns taxes on the corporate sale of energy commodities. In some Middle Eastern countries, the vast majority of state revenues are linked to the energy sector.

Commodity exporting countries have three key concerns regarding these tax revenues. First, the volatility of oil prices can create a volatile income stream for a sovereign state, which is clearly unwelcome, as government spending is much more stable than commodity prices. Second, it is unclear how long these commodity revenues will continue, as the commodities in the ground will not last forever. Specifically, there is a concern regarding **depletion**, which is the rate of extraction of a commodity relative to the remaining in-ground stocks. Third, governments would like to have a diversified economy, ideally earning tax revenues from other industries, such as technology, finance, and tourism, rather than relying almost exclusively on commodity revenues.

5.2 FOUR TYPES OF SOVEREIGN WEALTH FUNDS

This section discusses four types of sovereign wealth funds: stabilization funds, reserve funds, savings funds, and development funds.

5.2.1 Stabilization Funds

When an economy is building reserves, the initial store of wealth should be held in cash and liquid fixed income securities. In most countries, this wealth is held by the central bank and not segregated into a sovereign wealth fund. The central bank typically has three uses for official reserves: (1) implementing monetary policy, increasing and decreasing interest rates by trading fixed income securities, (2) intervening in the foreign exchange market, either to keep a currency at a fixed rate or to move a floating currency toward a desired level, and (3) injecting liquidity into the banking system to prevent crises. Most countries keep reserves in a central bank or other macroeconomic stabilization agency rather than separating these assets into a SWF charged with economic stabilization.

Countries with substantial commodity revenues have a much greater need for a stabilization fund, as commodity revenues are typically more volatile than the currency prices and interest rates typically of concern to central bankers. It is clearly not advisable for the government to have twice as much revenue during times of high commodity prices and half as much revenue during times of low commodity prices. It can be disruptive if economies substantially increase or decrease spending as commodity revenues change, so many governments seek to have relatively stable government spending. A sovereign wealth **stabilization fund** serves a countercyclical purpose through collecting excess commodity revenues during times of high commodity prices, and distributing saved wealth during times of low commodity prices. Alsweilem et al. (2013) discuss rules on the size of commodity revenues transferred from the general government budget into the stabilization fund. The easiest rule to implement and communicate is a fixed percentage, where the stabilization fund receives perhaps 20% of commodity revenues. Note that this is procyclical, where the government budget still receives higher revenues during times of higher commodity prices and lower revenues during times of lower commodity prices. Another rule is a hurdle price rule, where the stabilization fund receives all commodity revenue above a certain price, such as $50 per barrel. This will save the windfall revenues during the good times, but does not explicitly release stabilization funds into the government budget during times of low commodity prices. A final rule could be the deviation from a moving average rule, where commodity prices higher than a moving average rule move proceeds into the stabilization fund, while prices below a moving average move proceeds from the stabilization fund into the government budget.

With the decline in oil prices in 2015, there are concerns that many oil-based SWFs may need to use a large portion of their funds to augment government spending, reducing the assets of the SWFs. At the current trend, Kalter (2016) estimates that the following countries could exhaust their financial reserves within two to eight years: Iraq, Libya, Algeria, Saudi Arabia, Oman, Iran, Bahrain, and Yemen.

5.2.2 Reserve and Savings Funds

A sovereign state may wish to avoid the opportunity costs of investing entirely in cash and fixed income. Once the reserve adequacy standard has been met, a sovereign wealth fund is created with a total return objective. Al-Hassan et al. (2013) describe savings funds, pension reserve funds, and reserve investment funds. While many countries have both stabilization funds and reserve or savings funds, few countries have more than one of the following three types of funds.

Sovereign wealth **savings funds** are designed to bring intergenerational equity to a commodity-producing country by investing today's commodity revenues into a total return fund designed to benefit future generations. Similar to the university endowment funds described in Chapter 3, a spending rate can be instituted that provides income to future generations of beneficiaries while maintaining or growing the corpus of the fund. This is especially important in commodity-producing countries, as the commodity supplies may be depleted 20, 50, or 150 years from now and it is imperative that the country continues to benefit from commodity sales long after the in-ground supplies have been exhausted. Savings funds may be specifically invested in foreign currency assets designed to match the global distribution of where imported goods are purchased or the currency in which foreign debt is issued.

Many SWFs don't have clearly defined liabilities beyond a level spending rate. However, sovereign wealth **pension reserve funds** are designed to invest for high total returns in preparation for estimated future pension-like liabilities. In contrast to traditional pension funds that are funded with employer and/or employee contributions, pension reserve funds in the SWF category are funded with the proceeds of commodity or manufactured goods exports.

Sovereign wealth **reserve investment funds** are included in a country's reserve accounting as part of its reserves, but the funds invest in a total return portfolio in order to overcome the opportunity costs of the cash and fixed income dominated stabilization funds. Kunzel et al. (2010) notes that there are very few pension reserve and reserve investment funds, but those funds that do exist are very prominent (Korea Investment Corporation, China Investment Corporation, and Government of Singapore Investment Corporation).

5.2.3 Development Funds

Castelli and Scacciavillani (2012) describe sovereign wealth **development funds** as investment holding companies that have socio-economic objectives such as economic diversification, the development of strategic industries, or poverty alleviation. Development funds may focus on the domestic economy or choose to invest regionally or globally. Compared to reserve and savings funds, development funds may focus more on growing employment or diversifying the economy than on maximizing total returns. Development funds may be exclusively dedicated or largely allocated to alternative investments such as private equity and infrastructure.

Consider an economy where the majority of government revenue is derived from commodity exports. This economy is subject to volatile revenue streams as commodity prices and export volumes can vary dramatically over time. In addition, commodities are subject to depletion, where there is no guarantee how long the commodity reserves will continue to produce revenue. It is important, then, to diversify the economy beyond natural resources, such as Dubai has done by growing robust tourism and financial services industries. If and when the wells run dry, investments made by development funds will have ideally grown the economy to the point where it can be well supported by these non-commodity industries.

While savings and reserve funds seek to increase financial wealth that may be distributed in the future, development funds invest in infrastructure, human capital, and local industry to grow productivity and increase the future wealth of the country. These funds will typically be more concentrated, and may take on the flavor of private equity or even venture capital incubators. As concentrated investors in private firms, the return potential of development funds is substantial, as evidenced by the 16% annualized return of Temasek (a Singapore SWF) since its inception in 1974 through 2015.

Development funds bring a special set of conflicts to sovereign wealth funds. First, it is important to segregate the development fund from political influence and other government agencies. There can be a temptation for the governmental agency responsible for developing infrastructure to promote fewer projects as they are waiting for the SWF to develop projects that should be the responsibility of the infrastructure agency. Like any good private equity fund, projects or investments in a development fund should be evaluated on a financial basis, where the projects with the highest probability of success or the greatest profit potential should be financed. In some countries, there may be pressure to invest based on personal relationships that are more likely to enrich a corrupt friend of a politician than to create wealth or jobs for a country.

5.3 ESTABLISHMENT AND MANAGEMENT OF SOVEREIGN WEALTH FUNDS

This section discusses why sovereign wealth funds are established and how they are managed.

5.3.1 Four Common Motivations to Establishing a Sovereign Wealth Fund

There are four common motivations which may lead to the establishment of a SWF:[1]

1. Protect a nation's economy (and fiscal budget) from a potential decline or volatility in revenues;
2. Help monetary authorities to counter unwanted liquidity;
3. Grow the level of savings for future generations, particularly if the condition that drove the surplus is at a reasonable risk for depletion or reversal; and/or
4. Invest the money in infrastructure or economic growth projects today in order to strengthen a sector of the economy or grow a specific industry, especially to diversify away from commodity revenues.

5.3.2 Investment Management of Sovereign Wealth Funds

Exhibit 5.1 shows that the asset allocations of different types of sovereign wealth funds vary dramatically. While stabilization funds invest nearly exclusively in cash and fixed income, savings and reserve funds have highly diversified investments resembling the asset allocations of pensions, endowments, or foundations.

Exhibit 5.1 shows that savings funds, pension reserve funds, and reserve investment funds invest 72% to 74% in equities and other investments, which are largely alternative investments. Kalter (2016) notes that SWFs now have an estimated 40% allocation to alternative investments, similar in magnitude to that of endowments and foundations (43%), and family offices (44%), but far ahead of pension funds (17%).

Preqin (2016) shows that more than half of SWFs invest in public equity, fixed income, private equity, real estate, and infrastructure. A much smaller portion of SWF investors have made large allocations to hedge funds. Hentov (2015) estimates

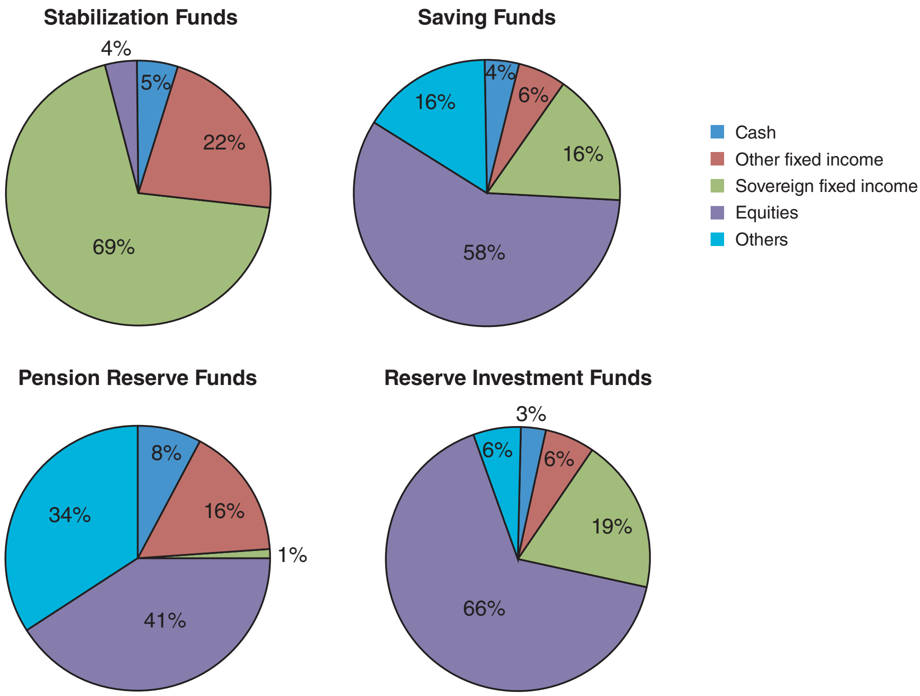


EXHIBIT 5.1 Asset Allocations at Sovereign Wealth Fund, by Type of Fund
Source: IMF, Global Financial Stability Report (2012), Al-Hassan et al. (2013).

that SWFs held $1.48 trillion in private market investments in 2014, an amount that could continue to grow as long as commodity and export based economies continue to earn large foreign currency revenues. The large size and long holding period of SWFs have led many investors to focus on direct investments rather than pooled investments, especially in infrastructure and real estate. Co-investments and joint ventures can reduce the fees of investing while the involvement of a general partner or joint venture partner can reduce the risk relative to a full direct investment program.

5.3.3 Dutch Disease and Sterilization Policies

The stabilization fund may be concerned about volatility in government revenues, as well as the value of the domestic currency. Consider a commodity producing country during a time of high commodity prices. The majority of the economy is focused on commodity exports, and the size of the economy is small relative to the large currency inflows from commodity sales. It can be disruptive if all of the (typically U.S. dollar) proceeds from commodity sales are converted into the local currency. There are two potential effects. First, inflation could result if the money supply increases faster than the availability of goods or investments in the local economy. Second, there is a fear of Dutch disease. **Dutch disease** occurs when large currency inflows (such as from the sale of large quantities of commodities) damages the long-run health of a country's other sources of economic prosperity (such as the country's manufacturing sector). In particular, the discovery of a tradable natural resource or a substantial increase in the price of such a resource has two major impacts. First, it will increase local wages by shifting workers from other sectors of the economy to the commodity sector. Second, the value of the local currency increases due to large inflows of cash. These two results reduce the competitiveness of the country's manufacturing sector, leading to its de-industrialization.

Many countries engage in sterilization. **Sterilization** is a macroeconomic policy in which a central bank or the government takes actions to counter the effects of an economic event (such as a commodity boom) and a balance of payments surplus on the country's economy. There are two types of sterilizations.

First, a country that is running a trade surplus may decide to intervene in the foreign exchange markets to prevent its currency from appreciating. In this case, the central bank will sell local currency and buy foreign currency in order to satisfy the increased demand for local currency by foreign importers of its products. If no further action is taken by the central bank, this will lead to inflation. To sterilize the local economy from the impact of the intervention in the foreign currency markets, the central bank will need to sell bonds to keep the money supply unchanged.

Second, a country may accumulate a significant amount of reserves not because of intervention in the foreign currency markets, but rather because of the revenues generated through the sales of natural resources that are completely or partially controlled by the government. If the government were to spend the revenues in the local economy, it could cause major disruptions in the local economy, especially if the local economy is small compared to the size of the inflows. In this case, the government may choose to invest much of the commodity proceeds outside of the home economy. For example, Norway may choose to invest much of its sovereign assets in dollar or euro denominated securities of other countries in order to not disrupt the local economy and to prevent the appreciation of the krone.

5.3.4 Managing the Size of a Sovereign Wealth Fund

After an extended period of high commodity prices, the size of a stabilization fund may be relatively large compared to the immediate spending and stabilization needs of the country. As the size of the stabilization fund grows, a question arises as to how large the stabilization fund should be. Given that the stabilization fund is invested almost exclusively in cash and liquid fixed income securities, there is a large **conservative investment opportunity cost**, where the longer conservative assets are left in cash, the greater the lost returns relative to what could have been earned on a balanced portfolio that includes investments in risky assets.

Reserve adequacy is the estimated size for stabilization fund reserves that is considered necessary before starting to invest in risky assets and moving funds into a total return portfolio. The necessary size of stabilization reserves are measured relative to the size of the potential shocks that can come from crises, whether from changes in commodity prices, trade flows, investment flows, or losses in the banking sector. Alsweilem et al. (2013) suggest that measures be developed relative to the size of external debt, imports, money supply, and the banking sector. Reserves may be held at the level of three to six months of imports, 20% of the money supply to assist the banking sector, while the Greenspan-Guidotti rule suggests that reserves be held against 100% of the value of all debts payable within twelve months. Alsweilem et al. (2013) suggest that SWF assets may continue to grow, as emerging markets countries have over $5 trillion in total reserves with needs of just $3 trillion to cover trade flows and money supply. Kalter (2016) estimates that over $18 trillion is currently held outside of SWFs, either in official foreign exchange reserves or other sovereign investment vehicles.

5.4 EMERGENCE OF SOVEREIGN WEALTH FUNDS

Carpantier and Vermeulen (2014) examined the period 1998–2008 and found evidence to support the theory that the surge in SWF establishments was determined by three main factors: (1) the existence of natural resource profits, and where SWFs tend to be established in countries that (2) run an autocratic regime and (3) have difficulties finding suitable opportunities for domestic investments. The net foreign asset position of a country was found not to be an explanatory variable, which indicates that the establishment of a SWF is unrelated to national accounting metrics.

Exhibit 5.2 illustrates the strong correlation over 1987–2015 that exists between the rate SWFs were established and the change in spot price of Brent Crude Oil, a common source of wealth for many of these nations. In particular, the total number of SWFs increased from 24 in 1998 to 62 by 2008 as the price for a barrel of crude jumped from $12.76 to $96.94 over the same period. As oil prices declined over

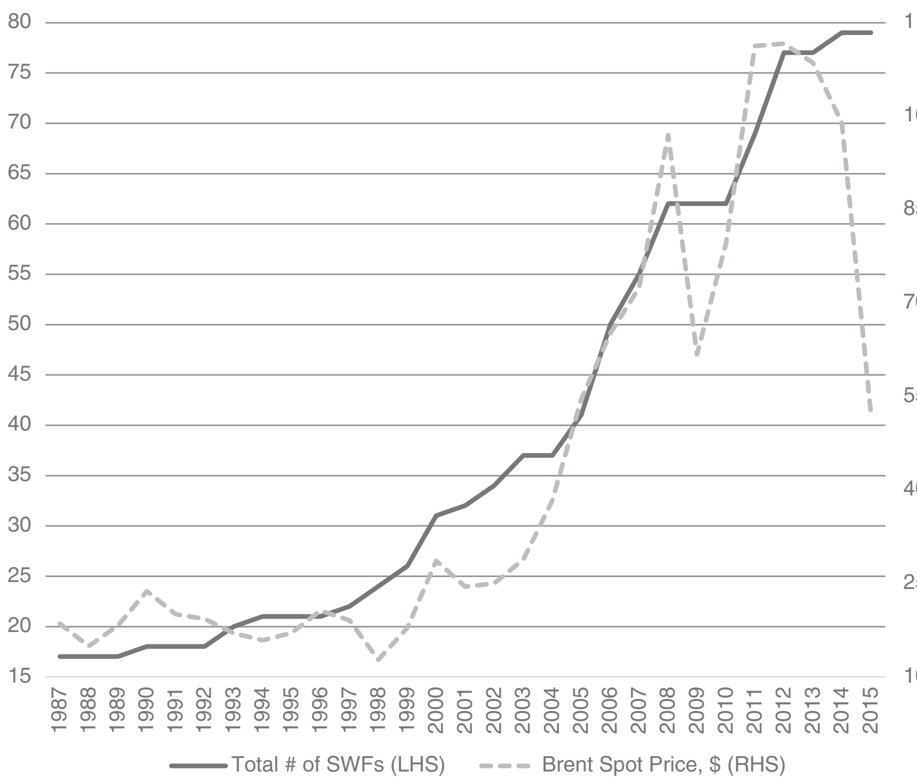


EXHIBIT 5.2 Correlating the Establishment of Sovereign Wealth Funds with the Price of Crude Oil
Sources: Sovereign Wealth Fund Institute, U.S. Energy Information Administration.

the subsequent two years, no additional SWFs were formed. However, as the price recovered between 2010 and 2012, 15 additional SWFs were formed, a 24% increase in the number of such funds. Only two SWFs were formed between 2012 and 2015, as oil prices had reached their peak in 2012 prior to declining dramatically over these three years. The relationship between the formation of SWFs and the price of crude is reminiscent of the strong correlation between company IPO issuance and the relative strength of their equity markets.

The largest SWFs are listed in Exhibit 5.3. Note that these 16 funds control over $6.1 trillion in AUM, while the full universe of over 70 SWFs have a total of

EXHIBIT 5.3 Assets Under Management at the World's Largest Sovereign Wealth Funds, 2015

	Commodity-Based SWFs ($Billion)	% GDP at PPP		Export-Based SWFs ($Billion)	% GDP at PPP
Norway—Government Pension Fund Global	$ 847.6	238.1%	China—China Investment Corporation	$ 746.7	3.9%
UAE—Abu Dhabi Investment Authority	$ 773.0	119.3%	China—SAFE Investment Company	$ 474.0	2.4%
Saudi Arabia—SAMA Foreign Holdings	$ 632.3	37.6%	Hong Kong—Hong Kong Monetary Authority Investment Portfolio	$ 442.4	2.3%
Kuwait—Kuwait Investment Authority	$ 592.0	205.6%	Singapore—Government of Singapore Investment Corporation	$ 344.0	72.9%
Qatar—Qatar Investment Authority	$ 256.0	80.0%	China—National Social Security Fund	$ 236.0	1.2%
UAE—Investment Corporation of Dubai	$ 183.0	28.2%	Singapore—Temasek Holdings	$ 193.6	41.0%
Saudi Arabia—Public Investment Fund	$ 160.0	9.5%	Australia—Australia Future Fund	$ 95.0	8.3%
UAE—Abu Dhabi Investment Council	$ 110.0	17.0%	South Korea—Korea Investment Corporation	$ 91.8	5.0%
	$3,553.9			$2,623.5	

Source: Sovereign Wealth Fund Institute, IMF, Author's calculations.

$7.2 trillion AUM, meaning that the assets of SWFs are highly concentrated, just as we saw with the university endowments in Chapter 3. This exhibit also shows that many countries also sponsor more than one SWF, with each managed to a separate objective as noted in Section 5.2.

5.5 GOVERNANCE AND POLITICAL RISKS OF SWFs

Many government funds have opaque investment policies and potentially substantial economic power from their large amounts of wealth. Uncertainty with regard to potential conflicts of interest can create costs that lower benefits to all parties. Transparency allows investors to better understand the impact that SWFs may have. Therefore, it may benefit SWFs to be transparent and ethical in their investments.

5.5.1 Governance of SWFs

Windfall commodity revenues in a country with tendencies toward corruption or underdeveloped financial systems must be treated with caution. There can be temptations to steal the money, invest in unwise domestic projects, or simply invest without the skill that is more common in markets with more developed financial markets.

SWFs should be independent of the government and the central bank, and managed in an external agency staffed with well-trained investment professionals. Bernstein et al. (2013) notes that when politicians are involved in a SWF, there is a greater allocation toward domestic investments, which is seen as a sign of poor decision-making. Within the SWF, well defined and transparent processes should be in place for investment management, risk management, and reporting of investment results. Efforts should be made to minimize the cost of an internally managed portfolio or to receive maximum value for the fees paid to external managers.

As will be seen later, Norway is the model of an independent, well-governed SWF with professional management, ultra-low investment costs, highly diversified portfolios, ethical investment policies, and well-developed transparency and governance policies. It is important to note that most of Norway's investments are made outside the country in minority stakes in publicly traded investments. This avoids the potential conflicts that can come with large domestic investments or with large stakes in foreign companies.

Perhaps the greatest concern about SWFs is the motives that strategic investors and their sponsoring governments may have beyond investment returns. Wagner (2013) states that the greatest political risk comes from countries with low levels of natural resources. SWFs in these countries may strategically invest in a way that seeks to ensure the availability of natural resources or may engage in technology transfer from firms in which they have made investments. Dyck and Morse (2011) discuss that SWF allocations are home-region biased and may focus on industries with a high potential for knowledge transfer, such as finance, transportation, energy, technology, and telecommunications.

Development funds may also be greeted with suspicion, as countries may engage in **protectionist policies** that favor domestic companies that are funded by the SWF over foreign companies in a given industry. Protectionism can take the form of quotas, tariffs, or governmental subsidies to domestic firms.

Finally, there is a question of the impact of SWF investments on their portfolio companies. Do the large investments of SWFs have a long-term impact on the stock price? Do SWF investors seek to control the actions of the portfolio companies in a way that would benefit the sponsoring government and their local economy? Avendaño and Santiso (2011) compare SWF holdings to those of private mutual funds and find that the investment decisions of SWF investors don't significantly differ from the decisions made by mutual fund managers. Furthermore, they did not find any evidence that SWF investment decisions are politically influenced, but rather that most SWFs go out of their way to not be perceived as controlling the interests of their portfolio companies. Finally, Megginson and Fotak (2015) cite a variety of studies that show inconclusive evidence regarding the impact of SWF investors on asset prices. Stocks added to SWF portfolios tend to outperform around the date of the acquisition announcement, but underperform over the following year, especially when compared to stocks purchased by other acquirers or professional investors. When SWFs make investments in distressed firms, it appears that the credit risk and CDS spreads of the firms decline, perhaps due to the assumption that a sovereign bailout would likely prevent the bankruptcy of these firms should greater distress arise.

5.5.2 Ten Principles of the Linaburg-Maduell Transparency Index

The Linaburg-Maduell Transparency Index, a method of rating SWF transparency, was developed at the Sovereign Wealth Fund Institute (SWFI) by Carl Linaburg and Michael Maduell. The index consists of the following 10 principles, where one point is assigned to a given fund for having satisfied its requirement. While the minimum rating is a 1, the SWFI recommends an index score of no less than an 8 (out of a maximum of 10) for a SWF to claim its transparency is adequate.

- Fund provides history including reason for creation, origins of wealth, and government ownership structure
- Fund provides up-to-date and independently audited annual reports
- Fund provides ownership percentage of company holdings, and geographic locations of holdings
- Fund provides total portfolio market value, returns, and management compensation
- Fund provides guidelines in reference to ethical standards, investment policies, and enforcement of guidelines
- Fund provides clear strategies and objectives
- Fund clearly identifies subsidiaries and contact information, if applicable
- Fund identifies external managers, if applicable
- Fund manages its own website
- Fund provides main office location address and contact information such as telephone and fax

5.5.3 Santiago Principles

Given the potential governance issues cited in Section 5.4.1, it becomes clear that it is important for SWFs to focus on good governance. In April 2009, the International

Working Group of SWFs set out the Santiago Principles, which define the generally accepted principles and practices (GAPP) of good governance of SWFs. Twenty-four governments have signed that they will seek to comply with the Santiago Principles, which are summarized in the following points:

- Governance and accountability.
- A clear objective.
- A sound legal framework.
- Adequate reporting systems.
- Integration in domestic policy formulation.
- Appropriate coordination.
- Clear rules on funding and withdrawal.
- Incorporating SWF data into macroeconomic data sets.
- Management of the nation's wealth.
- A clear investment policy.
- Diligence, prudence, and skill in investment practices.
- A robust risk management framework.
- Investment motivation.
- Public disclosure of policy purpose, governance framework, and relevant financial information.
- Refraining from pursuit of any objectives other than maximization of risk-adjusted financial returns.
- Public disclosures of general approach to voting and board representation.
- Fair competition in markets.
- Respecting and complying with all applicable host country rules, laws, and regulations.
- Not seeking advances of privileged information.
- Impact on global imbalances and capital movements.
- Disclosure of relevant financial information.
- Description of the use of leverage or disclosure of other measures of financial risk exposure.
- Execution of ownership rights consistent with the SWF's investment policy.
- A transparent and sound operational control and risk management system.

5.6 ANALYSIS OF THREE SOVEREIGN WEALTH FUNDS

This section takes a closer look at three SWFs in order to analyze and illustrate the economics of their management.

5.6.1 Government Pension Fund Global (Norway)

Alsweilem et al. (2013) explain that Norway had a major advantage by being a wealthy, diversified, developed economy before oil was discovered in the North Sea in 1969. Given that Norway didn't need the revenues generated from oil production, those funds were treated as a windfall subject to a large savings rate. Today, as much as 80% of the country's revenues are outside of the energy sector, so the role of the SWF to stabilize the economy from volatile oil prices is less important than in some Middle Eastern countries, where oil revenue can exceed 80% of state revenues. The Norwegian Government Pension Fund Global, then, can be categorized as a savings fund designed to allow future generations to benefit from this resource discovery.

All oil revenues are transferred to the SWF, while the fund transfers 4% of its value each year to the government budget. The government budget is designed to have a maximum non-oil structural deficit equal to the amount of projected draw from the SWF. That is, the government is able to balance its budget using the 4% income from the sovereign wealth fund. Now that the assets of the SWF exceed 230% of Norway's GDP at purchasing power parity, the 4% draw from the SWF is over 9% of GDP and an even larger portion of government spending. If the fund is able to earn a 4% real return, the real value of the fund can be maintained in perpetuity, even after the oil supplies have been exhausted. After the discovery of oil in the North Sea in 1969, oil production rose steadily, peaking around the year 2000. While gas production has been increasing in recent years, oil production has declined by nearly 50% in the last fifteen years, showing that oil supplies may be nearly exhausted in the next twenty years in the absence of new discoveries. After oil revenues cease to flow into the fund, it is imperative that the SWF be managed as a perpetual endowment, as the windfall experienced over the last 50 years in unlikely to be repeated, so future generations will rely on the savings of the SWF to benefit from this large oil discovery. Norway currently does not have large investments in inflation-hedging assets, as oil and gas reserves largely serve that purpose. However, as oil and gas reserves are depleted, it would be expected that investments in real assets would increase to serve the long-term inflation hedging role.

Chambers, Dimson, and Ilmanen (2012) contrast the Norway model to the endowment model or Yale model described in Chapter 3. In the **Norway model**, asset allocations are dominated by highly diversified liquid assets that are managed at extremely low costs, which seem to be the opposite of the Yale model, which maximizes exposure to illiquid assets invested with external managers. Norway invests around 60% in global equities, 35–40% in global fixed income, and is slowly investing to fill a recently added 5% allocation to real estate.

With a small staff of less than 500 employees and a large fund, the strategy is aimed at passive management, acknowledging that active management fees and the market impact of such a large fund can impair returns. Norway has a very large and diversified fund, with holdings in more than 9,100 equity securities, including owning more than 2% of the outstanding shares of over 1,200 companies.[2] The scale of the fund is massive, with the SWF owning 2.4% of the market capitalization of the European equity market, 1.4% of the value of emerging and frontier equity markets, and 1.3% of the world's other developed markets. Similarly, the fund is invested in the debt securities of over 1,100 issuers in 31 currencies. The assets are largely internally managed at a cost of just 5.9 basis points.

Norway is seen as a leader in SWF transparency and governance as well as in its stance on social investing. The SWF publishes a 100-page annual report in addition to a separate report on responsible investment. In the annual report, detailed asset allocations are presented, including fund and benchmark returns for the entire fund as well as for subportfolios, such as European equities. Risk management policies are disclosed, as are specific breaches of risk management policies or operational risk events during each year. Compensation levels of top employees are disclosed, as is the structure of the performance-based pay program offered to nearly 200 employees. The rules for contributions to, and distributions from, the fund are clearly set and are stable over time.

The Norwegian government maintains a list of prohibited investments[3] that exclude a number of equity investments from the SWF portfolio. These exclusions ban investments in a number of industries, including the production or use of coal, certain types of weapons, and tobacco. Company-specific inclusions are listed for firms that are determined to have damaged the environment, violated human rights, or have issues with corruption, ethics, or conflict zones. The fund specifically targets investments in environmental technology firms and regularly votes at equity shareholder meetings.

5.6.2 China Investment Corporation

In June of 2015, China was reported to have held over $4 trillion in foreign exchange reserves, most of which was earned through the export of manufactured goods and a currency that was undervalued for an extended period of time. As can be seen in Exhibit 5.3, China and Hong Kong have over $1.8 trillion in four large funds, while the balance of the reserves are likely held by the central bank. China's State Administration of Foreign Exchange (SAFE) is located within the central bank and functions as a stabilization fund, with significant infusions into the state-owned banking. Due to the size of SAFE, the fund has started to invest in risky assets in the pursuit of higher returns than earned with the previous focus on cash and fixed income investments. The Chinese funds operate quite differently than Norway, especially SAFE, which is not known for its transparency.

Alsweilem et al. (2013) explain that China Investment Corporation (CIC) was established in 2007 in order to increase independence of the investment arm from the central bank. CIC can be classified as a reserve investment fund, which seeks higher returns. While many SWFs focus investments outside of their home countries, CIC controls Central Huijin, a holding company originally started by SAFE's injections into the banking system, which owns between 35% and 66% of five state-owned Chinese banks. Other divisions of CIC are CIC International and CIC Capital.

While Norway is highly diversified and highly liquid, CIC has more concentrated investments, a large allocation to private equity, and significant exposure to external management, which is more similar to the Yale model. As of the end of 2014, the asset allocation of CIC is approximately 44% public equity, 26% long-term investments, 15% fixed income, 12% absolute return, and 3% cash. Over 70% of the fixed income portfolio is invested in sovereign debt, while over 22% of the equity portfolio is held in the finance sector.

Global leaders have expressed concern over the political influence of CIC, especially when investing in specific industries, such as oil and gas.[4] Alsweilem et al. (2013) report that CIC invested in distressed financial firms in 2008, including $3 billion in the Blackstone Group and $5.6 billion in Morgan Stanley. Those firms were later hired by CIC to advise on $500 million in hedge funds and $800 million in real estate investments, respectively. With the size of these stakes, CIC is not a passive investor, but has taken seats on a number of corporate boards.

CIC also invests heavily in real assets, especially those that are seen as providing inflation protection, or hedges against a falling U.S. dollar, or those improving the resource security of the Chinese population. The allocation of CIC has been moving away from public equities and toward greater allocations in private equity and direct investments, especially in oil and gas, mining, utilities, and infrastructure. CIC includes these assets in the long-term investments category. CIC started CIC Capital in 2015, which has specific goals to oversee private investments in infrastructure and farmland.

5.6.3 Temasek Holdings (Singapore)

Like China, Singapore sponsors a number of funds, each with a different objective. Singapore's Government Investment Corporation (GIC) functions as a reserve investment fund, seeking to maximize returns on a global portfolio, with over 80% invested in a diversified portfolio of liquid equities and fixed income securities. Singapore's Temasek Holdings is an example of a development fund that has been funded with the proceeds of privatizations, which are the sale of government owned assets to private investors. While a primary goal of Temasek may be the growth of employment and economic development in Asia, a portion of the returns from Temasek's investments are paid as dividends to the government of Singapore.

As a development fund, Temasek makes large strategic investments to develop certain sectors. Singapore is a wealthy, but small, city state, which has developed its wealth through shipping, engineering, and financial services. Given its small size and high population density of 5.5 million people in just 719 square kilometers, Singapore focuses investments on areas driven more by intellectual capital than by natural resources or heavy manufacturing.

Temasek's portfolio is highly concentrated, with 28% of holdings in Singapore, and 42% in the rest of Asia. At the end of 2015, over two-thirds of the portfolio was invested in financial services, transportation, industrials, as well as telecommunications, media, and technology (TMT). Temasek's portfolio focuses on public and private equity investments, with one-third each invested in unlisted assets, listed large blocks above a 20% holding, and liquid investments with less than a 20% stake in the underlying firms. Temasek owns 1%–3% stakes of some large firms, such as Industrial and Commercial Bank of China, Gilead Sciences, Alibaba Group, and Ping An Insurance Company of China. While most SWFs take small stakes in firms, Temasek controls the majority of shares in firms such as Singapore Airlines, Singapore Telecommunications, Neptune Orient Lines, and SMRT Corporation, while owning all of the equity of Singapore Technologies Telemedia, MediaCorp, PSA International, Singapore Power, and Pavilion Energy.

5.7 CONCLUSION

Over the past 10 years, SWFs have increased their assets by over $4 trillion, including asset returns and the establishment of new funds from the proceeds of commodity and merchandise exports as well as privatizations. The initial allocation of a new SWF is often close to 100% in cash and fixed income, as the assets are used to stabilize the economy against volatility in export revenues, tax revenues, and foreign trade and capital flows. Once a central bank determines that the level of reserves is adequate for the purpose of economic stabilization, assets are moved into a vehicle seeking higher returns. As the AUM of a SWF rises and the sophistication of the investment team grows, allocations to alternative investments tend to increase. To date, these investments have been focused on private equity and real assets with lesser interest in hedge funds.

NOTES

1. Please see Rozanov (2005).
2. 2014 Norwegian Government Pension Fund Global Annual Report, www.rijpm.com/pre_reading_files/Eivind_Oy_Norwegian_MinFin_June_2014.pdf.
3. www.nbim.no/en/responsibility/exclusion-of-companies/.
4. Koch-Weser, I., and O. Haacke. June 2013. China Investment Corporation: Recent Developments in Performance, Strategy, and Governance. U.S.-China Economic and Security Review Commission.

REFERENCES

Al-Hassan, A., M. Papaioannou, M. Skancke, and C. Chih Sung. 2013. "Sovereign Wealth Funds: Aspects of Governance Structures and Investment Management." Working paper No. 13/231, International Monetary Fund, November 13. Available at https://www.imf.org/external/pubs/cat/longres.aspx?sk=41046.0.

Alsweilem, K., A. Cummine, M. Rietveld, and K. Tweedie. 2013. "Sovereign Investor Models: Institutions and Policies for Managing Sovereign Wealth." Working paper, Harvard Kennedy School.

Avendaño, R., and J. Santiso. 2011. "Are Sovereign Wealth Funds Politically Biased? A Comparison with Other Institutional Investors." In B. Narjess and C. Jean-Claude (eds), *Institutional Investors in Global Capital Markets* (pp. 313–353). Bradford, UK: Emerald Group Publishing Limited.

Bernstein, S., J. Lerner, and A. Schoar. 2013. "The Investment Strategies of Sovereign Wealth Funds." *Journal of Economic Perspectives* 27:219–238.

Carpantier, J., and W. Vermeulen. 2014. "Emergence of Sovereign Wealth Funds." University of Oxford, OxCarre Research Paper 148, November 16. Available at www.economics.ox.ac.uk/materials/papers/13842/paper148.pdf.

Castelli, M. and F. Scacciavillani. 2012. *The New Economics of Sovereign Wealth Funds*. Chichester, UK: John Wiley & Sons.

Chambers, D., E. Dimson, and A. Ilmanen. 2011. "The Norway Model." *Journal of Portfolio Management* 38, no. 2 (2012): 67–81. Available at http://ssrn.com/abstract=1936806 or doi:10.2139/ssrn.1936806.

Dixon, A., and A. Monk. 2011. "The Design and Governance of Sovereign Wealth Funds: Principles & Practices for Resource Revenue Management." October 30. Available at http://ssrn.com/abstract=1951573 or doi:10.2139/ssrn.1951573.

Dyck, A. I. J., and A. Morse. 2011. "Sovereign Wealth Fund Portfolios." Working paper, Chicago Booth School of Business.

Gwartney, J., R. Stroup, R. Sobel, and D. Macpherson. 2003. *Economics: Public and Private Choice*. 10th Edition. Mason, OH: South-Western College Publishing.

Hentov, E. 2015. "How do Sovereign Wealth Funds Invest? A Glance at Sovereign Wealth Fund Asset Allocation." State Street Global Advisors, December.

International Monetary Fund. 2014. "Sovereign Asset-Liability Management." IMF policy paper, June 10. Available at https://www.imf.org/external/np/pp/eng/2014/061014.pdf.

International Working Group of Sovereign Wealth Funds. 2008. "Generally Accepted Principles and Practices (GAPP)—Santiago Principles," October. Available at www.iwg-swf.org/pubs/eng/santiagoprinciples.pdf.

Kalter, E. 2016. "Sovereign Wealth Fund Investment Trends." Tufts University.

Kunzel, P., Y. Lu, I. Petrova, and J. Pihlman. 2010. "Investment Objectives of Sovereign Wealth Funds—A Shifting Paradigm." Working paper, IMF.

Megginson, W. and V. Fotak. 2015. "Rise of the Fiduciary State: A Survey of Sovereign Wealth Fund Research." *Journal of Economic Surveys* 29(4):733–778.

Preqin. 2016. "2015 Preqin Sovereign Wealth Fund Review."

Rozanov, A. 2005. "Who Holds the Wealth of Nations?" State Street Global Advisors' Official Institutions Group, August. Available at http://piketty.pse.ens.fr/files/capital21c/xls/RawDataFiles/WealthReportsEtc/SovereignFunds/General/Rozanov2005.pdf.

Wagner, D. 2013. "SWFs: Investment Objectives and Asset Allocation Strategies." Working paper, Justus-Liebig-University Giessen.

CHAPTER 6

The Family Office Model

A family office refers to an organization dedicated to the management of a pool of capital owned by a wealthy individual or group of individuals. In effect it is a private wealth advisory firm established by an ultra-high-net-worth individual (UHNWI). This chapter discusses the family office model, including the types of family offices, managing wealth and relationships across multiple generations of a family, taxes, lifestyle investments, and philanthropic activities.

6.1 IDENTIFYING FAMILY OFFICES

Information on the growth of family offices is limited because of the highly confidential and private nature of family offices and the wealthy individuals they serve. As of 2013, the number of family offices in the United States was approximately 3,000 while globally the number was close to 5,000. In total, these offices manage approximately $3 trillion.[1]

In some cases the family office is referred to a "single" family office, and this means that the management of capital and financial affairs is dedicated to a single individual or a single family. Some offices are set up to serve multiple families—a multifamily office. This may be done to share operating expenses and back office administration, or to collectively pool investment ideas.

As a general rule of thumb, dedicated family offices are considered operational at $1 billion of assets under management (AUM). For amounts smaller than that size, it would make sense for the wealthy family to join a multifamily office to share expenses and achieve economies of scale. It is estimated that running a single family office costs approximately 60 basis points per year.[2]

The source of income for a family office can be as varied as the underlying family that it serves. In some cases, the capital is spun off from an operating company (e.g., Mark Zuckerberg of Facebook). In other cases, the family office might be funded with what is known as "legacy wealth." This refers to a second, third, or later generation of family members who have inherited their wealth from a prior source of capital generation. The family office can be funded with cash, real estate, natural resources, equity securities, or even art.

6.2 GOALS, BENEFITS, AND BUSINESS MODELS OF FAMILY OFFICES

Family offices are diverse with respect to their goals, benefits, and business models.

6.2.1 General Goals of the Family Office

The goals of the family office vary dramatically. One goal can be preserving wealth for the next generation. Another goal can be to give away all of the wealth. A third goal can be to place the family wealth into a foundation for long-term charitable goals. A more immediate goal might be to maintain a certain lifestyle. This can include funding purchases like homes, art, automobiles, private jets, and the like.

Another survey of single family offices found wealth management to be the most common goal for the family. Wealth management and accounting consolidation seem to be very important goals for every family office. More simply, a key benefit to a family office is to have a consolidated function for family office wealth management, accounting, and control.

The benchmark returns for the family offices are also very different. Some family offices are benchmarked to an absolute return hurdle: Treasury bills or LIBOR + X% or an inflation target. Some families have benchmarks defined in terms of objectives instead of a certain percentage. For example, the benchmark for one family office is simple and clear on its face: Don't lose money. Not surprisingly, this family office has a very conservative asset allocation that we will observe later in this chapter.

6.2.2 Benefits of the Family Office

A family office provides several benefits over a private bank or traditional asset manager. A family office serves as the central source for financial management for a family or group of families. As a result, it is a key source of information and advice on all family financial affairs. Family offices also have the benefit of absolute privacy. Further, the family office is intimately aware of the personalities, risk tolerances, individual goals, and the spending needs of the family members. The staff of the family office is dedicated exclusively to the financial and philanthropic goals of the family and its members. Such dedicated advice and counsel are hard to duplicate outside of a family office.

Exhibit 6.1 shows what benefits or functions matter most to ultra high net worth individuals (UHNWIs) in setting up a family office.[3] Clearly, wealth management ranks highest. But also important is conflict-free advice—the benefit of a dedicated staff, confidentiality, and sophisticated investment advice. Further down the ranking are the softer benefits such as education of family members, philanthropy, and concierge services.

Another key benefit is keeping the family money consolidated in one central core. This can become more difficult as more generations emerge with their individual goals and objectives. It is typically not an easy task to get family members to agree on matters related to money.

6.2.3 Models and Structure of the Family Office

Family offices manage capital for individuals whose goals are as varied as the colors in the rainbow. Consequently, there is no "family office model" as a counterpart to the "endowment model" of universities. Each family office follows a unique game plan that reflects the individual members of that family, their specific goals, the source of their wealth, and their lifestyle needs.

Exhibit 6.2 presents the business models for several UHNW family offices.[4] First, notice that in some cases the founding family is involved in the management of the

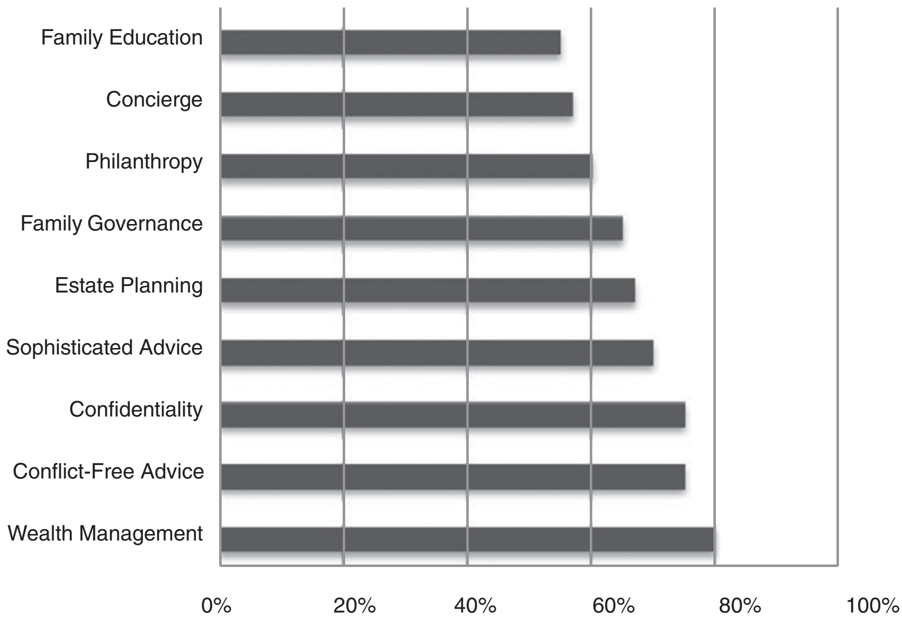


EXHIBIT 6.1 Benefits of a Family Office

family office. Sometimes, family members even make investments on behalf of the family. There is no right or wrong here; it really is the individual nature of the family office. In some cases, the second and third generation of family wealth may not have day jobs and may want to be actively involved in the investing of the family wealth. This can be a good way to learn about and help construct the overall portfolio for the family. The downside is that these family members may not have rigorous training or experience in asset allocation and portfolio management.

One dimension along which family offices differ is whether they manage outside capital for external clients. Previously, in the United States, family offices were allowed to operate under a "safe harbor" with the Securities and Exchange Commission (SEC) that allowed them to accept up to 15 outside clients. With the 2010 passage of the Dodd-Frank Act by the U.S. government, this safe harbor has been eliminated from the U.S. securities laws. Therefore, if a family office accepts \$1 from a non-family member, it may be required to register with the SEC. A family member is defined as any member of the direct bloodline of the founder of the family office plus spouses of family members. Anyone outside that family circle is considered an outside client and brings with it the full burden of SEC registration, regulatory adherence, and government agency scrutiny.

Family lines can become a bit murky regarding who, exactly, is a family member. Currently, the SEC allows family offices to advise up to 10 generations of family (and spouses) related to a common designated ancestor. However, in a recent well-publicized case, the SEC allowed the movie star Meryl Streep to keep her money with a family office even though she had divorced the family member from whom she gained access to the family office in the first place.[5]

Once we can define who is not a family member, the regulatory burden becomes significant. By accepting outside clients, the family office must develop an internal

EXHIBIT 6.2 Various Models of a Family Office

	Estimated AUM	Family Actively Engaged	Size of Team	Structure of Team	Benchmark	In/Outsource Infrastructure	Hedge Funds	Direct Private Equity	External PE Firms	Real Assets	Internal Management	Asset Allocation	Accept Outside Clients
Family Office A	$14 billion	Yes—in VC	25	7 in VC, 5 in PE, 5 in public markets, 4 in credit, 4 for risk and asset allocation, rest as analysts	Outperform the S&P 500	In-house—team size undisclosed	Only 10 to 15	Yes, $50 to $200 million	10 external firms	Minimal—not their strength	Yes, direct VC investing	No formal risk budget or allocation	Yes, in their Capital Funds. Also willing to co-invest.
Family Office B	$15+ billion	No	25	By asset class	Would like to earn 10% across total portfolio	Full back and middle office In-house—team of 50—full bookkeeping to performance attribution	None—they have their own L/S and macro trading in-house	Yes, $50 million and up	8 to 10 PE funds: U.S. mid market, distressed debt, RE	Yes, but rely more on outside managers	Yes, concentrated long only, L/S equity, bonds, currency	Quarterly risk and asset allocation meetings—can go to zero in any asset class	No, but will co-invest in other direct deals
Family Office C	$7–$8 billion	No	20	2 to 4 persons per asset class team; 3 persons asset allocation team, 3 persons operational due diligence team	Not disclosed: T-bills + a premium	In-house—12 people who handle everything on-site—they like the immediate interaction	Yes, 90% of the capital is allocated externally to hedge funds	Very limited	Only 2	Yes—allocated to external managers	None	Formerly meet with the Foundation 3 times a year	No, but willing to co-invest with other family offices
Family Office D	$5+ billion	Yes	9	1 direct deal guy, 2 structured credit, 2 traders, 2 for public equities, CIO for asset allocation, CEO makes macro bets	T-bills plus a premium	Outsource all back and middle office	Run their own internal discretionary macro portfolio plus external hedge funds to diversify	Yes, but very limited.	None yet	Don't invest in real assets: "not our expertise"	Very active internal management across equities, bonds, credit, currency, commodities	Left to the CIO but no formal meetings	No, but willing to co-invest with other family offices

Family Office E	$6+ billion	Yes	10	Across traditional asset classes	T-bills + 4% to 5%	Hybrid between Viteos system for accounting and in-house for capital calls	Yes, portfolio is geared more toward absolute returns	No	Very limited	Yes	Focus on internal bond management: investment-grade, high-yield, short-term cash	Different allocations for different family members	Looking to accept 2 to 4 other families to spread costs and create economies of scale
Multi-Family Office F	$25–$250 million	No	22	Organized by asset class	Blended based on asset allocation	Outsource to custodian	Yes—significant source of their alpha	No	Only 3 firms	Commodities but limited RE	None	Formal asset allocation committee once a year	Yes, 20 clients, and 8 client PMs to work on asset allocation
Family Office G	$3.3 billion	No	8	4 accountants, 4 investors, including CIO	None. Goal is not to lose money.	Use their external bank for performance, custody, risk management	Yes—a significant commitment to hedge funds. No long-only public equity.	No—consider their investments in the family operating company to be similar to direct PE	Very small. They consider the family operating companies to be their equivalent of PE	Some energy exposure through Riverstone PE fund—very small	None. All outsourced primarily to HF and stock and bond managers	No formal asset allocation model. Meet quarterly.	No, they have several family members that keep them fully occupied.
Family Office H	$3+ billion	Yes	7	4 investment, 2 accountants, 1 support staff	Inflation	Outsource risk management	A blend of HF managers but tilted toward those strategies with less market risk	None	Very limited	Oil and gas	None, all external managers	No formal asset allocation or committee	No, regulatory burden is simply too high.

compliance department to ensure that all external clients are treated on an equal footing with the family members. In other words, the sponsoring family cannot be selective in choosing the best investment ideas just for their family; they must share all investments equally with their outside clients. This can be an especially prickly situation with family members who source the good investment ideas in the first place. They may not want to share their investment insights with outside clients who contribute nothing to the generation of investment ideas. In addition, the family office must develop a sophisticated reporting infrastructure to provide monthly and quarterly reports to the external clients. Last, there are ongoing reporting requirements to the SEC and other regulators that can require a small army of lawyers, accountants, and administrators to fulfill.

The key point is that if a family office intends to accept outside clients, it must do so with the full knowledge that this is a very expensive proposition. One family office estimates that its operating costs more than doubled once it accepted outside clients. Other family offices—notably that of George Soros—immediately shut down their outside investment activities and returned capital to all of their external clients when the SEC safe harbor expired. Still other family offices—Rockefeller & Co., for example—continue to accept outside clients and have become fully registered with the SEC as investment advisers.

6.3 FAMILY OFFICE GOALS BY GENERATIONS

The number of generations has a substantial impact on asset allocation and risk taking.

6.3.1 First-Generation Wealth

The first generation of wealth creation (new money) is more likely to be concerned with wealth preservation—keeping the newly created wealth intact. It is successive generations (old money) that turn to generating wealth growth with an established pool of assets.

However, family offices are more likely to be set up past the first generation. Exhibit 6.3 demonstrates the generational transfer of wealth and when it comes under the management of a family office.[6] This exhibit demonstrates that family offices are usually established after the first generation of wealth is created.

This may not be as surprising as it looks. The first generation of family wealth is usually tied to an operating company. As a result, the first generation is typically focused on creating the wealth, not managing it. It is only after the wealth has been created and typically locked up in a transaction of some kind (e.g., the sale of an operating company) that the family begins to think about the ongoing maintenance of the wealth.

6.3.2 Risk Management of First-Generation Wealth

There are special concerns that arise from **concentrated wealth**, which occurs when the vast majority of the assets are poorly diversified such as being held in a single company. Concentrated wealth positions typically arise because the newly wealthy often

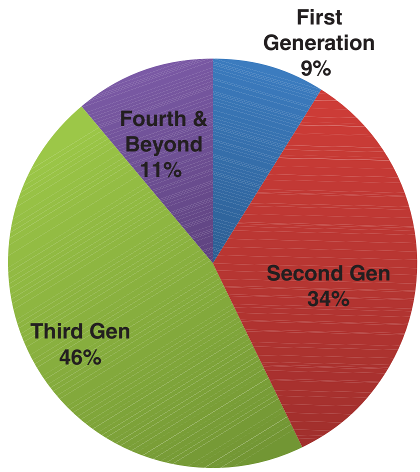


EXHIBIT 6.3 Generations Served by a Single Family Office

earn their fortune from a privately held family business or serving as an entrepreneur or chief executive of a public company. While starting or managing a public company can create substantial wealth, there are also regulations and shareholders to respect. Many executives of public companies are required to refrain from trading their shares during times when they are in possession of material nonpublic information, such as around announcements of corporate profits or merger transactions. Purchases or sales of company stock may also be required to be reported to regulators and/or shareholders. Both the disclosure requirement and the restricted holding periods serve to make the company stock less liquid for the executives at the company than for other shareholders. Sales of stock may incur a substantial tax liability, as they are often held at a large increase in value relative to the acquisition or granting price.

Many chief executives regularly invest the vast majority of their wealth in the stock of the company they lead. This may be a sign of confidence or solidarity with shareholders or the result of receiving stock options or grants that may require some price move or the passage of time before the sale is allowed. These executives may be interested in regularly scheduled sales of stock, perhaps 1% of holdings sold immediately after the release of each quarterly earnings report, in order to diversify their portfolios and grow assets outside of the company over time. The proceeds of these sales may be invested in a completion portfolio. A **completion portfolio** is a collection of assets that is managed with the objective of diversifying and managing the aggregated risks of the concentrated portfolio and the completion portfolio. The assets purchased should have a low correlation to the assets held in the company stock portfolio (the concentrated portfolio). For example, an entrepreneur who owns a substantial stake in a U.S.-based technology company would not buy any technology stocks, but focus purchases on sovereign debt, real assets, and foreign stocks, perhaps in the finance or health care sector, resulting in an increasingly diversified portfolio, even when including the concentrated stock holding.

Some executives may hedge their concentrated holdings with long-term stock options, such as through an options collar, which sells call options and buys put options. This strategy delays realized gains for tax purposes, while limiting the potential gains and reducing the potential losses relative to the current market price. Shares can also be held as collateral for loans, which can provide liquidity, but can also multiply losses should the value of the pledged shares decline below the value of the loan.

Generating wealth through a privately held business has even more extreme illiquidity than serving as the executive of a public company. Much of the wealth that is managed by family offices is generated by a **liquidity event** such as the sale of the family business in a merger transaction or in an initial public offering. These liquidity events may generate $10 million or more in proceeds in a single week, dramatically changing the wealth management strategies of the family. Before the liquidity event, nearly all of the family wealth is held in a private business that is highly illiquid and difficult to value. After the liquidity event, the family has a large amount of cash that can be easily invested in a diversified portfolio. Some advisers encourage sellers of private businesses to take some time to invest the proceeds of their life's work, perhaps a year or more, especially during times of overvaluation in the equity markets. This is the time for education on investment markets and reflection on family goals, including the distribution of the wealth and whether the goal is to preserve or multiply the wealth. The time can also be used to have discussions with attorneys and financial advisers as well as consider whether to start a single family office or join a multifamily office.

6.3.3 Benchmarking First-Generation Wealth

The T-bills/LIBOR + X% benchmarks reflect one style of family office management typically associated with first-generation wealth. Often the initial wealth is created, or continues to be created, by an operating company. The operating company, in many cases, produces a large concentration of wealth from a single source. Using principles of portfolio management, there is considerable risk when a significant portion of wealth is concentrated in a single asset.

When this is the case, many family offices take more of an absolute return approach to their asset allocation. The family office portfolio is used to diversify the single company risk of the initial or ongoing wealth generation. Under this approach, more of the portfolio is geared toward hedge funds, credit, fixed income, relative value, and other assets that are less geared toward market risk.

In fact, when you examine those family offices that have an absolute return benchmark (like T-bills + X%), the allocation to private equity is very limited. Again, this points toward taking less market risk in the family office to balance the risk contained in the family operating company. In fact, one family office considers the ongoing ownership of the family operating company to be its private equity portfolio.

6.3.4 Goals of the Second Generation and Beyond

Taxes play an important role in portfolio construction and asset allocation within family offices, particularly in relation to second- and third-generation wealth. At this stage of the family life cycle, the initial wealth has been created from an original underlying source—usually an operating company that has been sold. In this situation, the family office portfolio becomes the new source of wealth generation.

When constructing an optimal portfolio for a family office and building an efficient frontier, it is important to use after-tax returns. At this generation of the family life cycle, the family office portfolio is geared more toward growth assets—those assets that generate long-term capital gains instead of assets that are more short-term oriented either in their income generation or capital gains.

Consider Exhibit 6.4. This shows the asset allocation of a well-known university endowment and that for an UHNWI. One observation that can be quickly made is

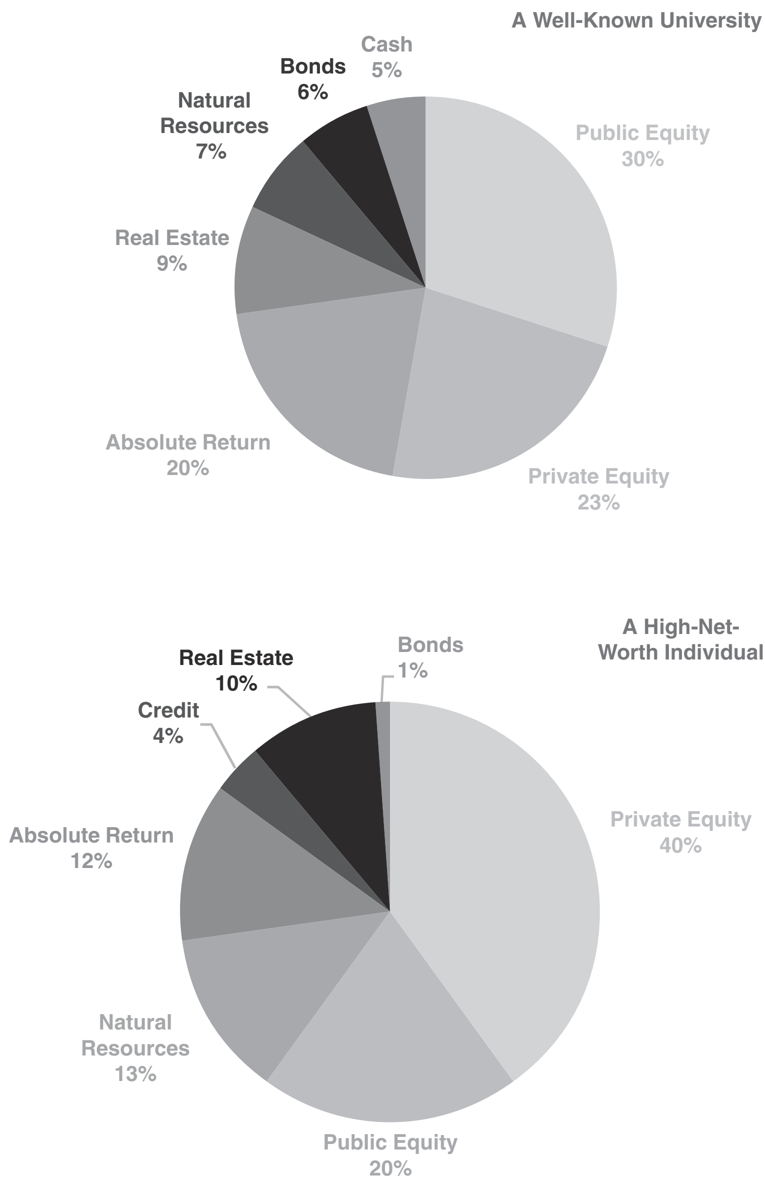


EXHIBIT 6.4 Sample Asset Allocation for Endowments versus Family Offices

the much larger allocation to private equity for the individual versus the university. Endowment portfolios are usually considered to have an aggressive risk profile with a considerable portion of the endowment allocated to risky assets. However, family office portfolios take this a step further.

Reviewing Exhibit 6.4, one can see that 87% of the family office portfolio is allocated to long-term risk-based assets that generate long-term capital gains: private equity, public equity, natural resources, credit, and real estate. By contrast, the university endowment has only 69% of its portfolio allocated to these long-term assets. Not surprisingly, the endowment has a much larger allocation to bonds and cash and absolute return investments (31%) than the family office (13%).

The higher allocation to risky assets by the family office serves two goals: (1) tax efficiency and (2) wealth generation. Again, keep in mind that Exhibit 6.4 demonstrates the portfolio allocation for a family office that has already established its wealth from an initial source and now uses the family office portfolio as the driver of wealth growth.

Exhibit 6.4 raises another issue for a family office compared to an endowment. The higher allocation to private assets that generate long-term capital gains can result in a very illiquid portfolio. Exhibit 6.5 demonstrates this problem.

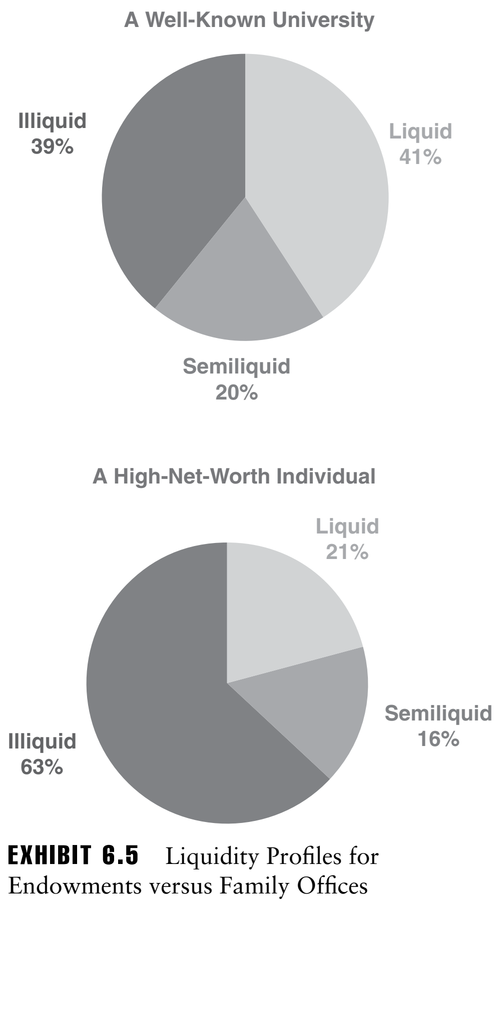


EXHIBIT 6.5 Liquidity Profiles for Endowments versus Family Offices

Again, we compare the well-known university to the UHNWI, but this time along the dimensions of liquidity. Illiquid assets are defined as those assets that cannot be sold within one year without selling at a discount. Liquid assets are those assets that can be sold in less than three months without selling at a discount, and semiliquid assets fall into the three-month to one-year time frame. In reality, every asset can be sold within a year—a home, a private equity portfolio, a ranch. The problem is that chunky/illiquid assets generally must be sold at a discount if the UHNWI wants liquidity quickly.

Looking at Exhibit 6.5, we can see how lopsided the UHNWI portfolio is with respect to illiquid assets compared to the university endowment. The university has achieved an almost perfect balance between its liquid and illiquid assets, with semiliquid assets filling the gap between the other two categories. However, for the UHNWI, illiquid assets represent 63% of the total portfolio. For this reason, many UHNWIs have a credit line or liquidity backstop negotiated with an external bank. Yet, this liquidity profile demonstrates an advantage of the family office over the endowment. Unlike an endowment, the family office does not have a university operating budget to backfill each year. To the extent that the family has a low spending rate, the family office can capture even more of the liquidity premium than sophisticated endowment offices can. If there is a desire for the family to sustain a high spending rate, the allocation to illiquid assets will be lower.

6.4 MACROECONOMIC EXPOSURES OF FAMILY OFFICES

Another difference between family offices and endowments is a concern for the overall macroeconomic variables that can affect an investment portfolio. Endowments, for example, are much more focused on beta drivers—the systematic risk premiums attached to asset classes. Beta drivers are effective tools for asset allocation, risk budgeting, and income generation. Exhibit 6.6 shows the beta drivers for both an endowment and a UHNWI.

However, there is a growing trend among wealthy families to better understand how their wealth is tied to global macroeconomic factors. The financial crisis of 2008 demonstrated to investors large and small that many asset classes became highly correlated with one another, resulting in significant declines in asset values. Macroeconomic factors are a way to look beyond asset classes and beta drivers to better understand the true economic forces that can impact a portfolio.

Exhibit 6.6 shows a simple diagram of how each asset class can be broken up into global economic factors. Each asset class can be shown to be a combination of four macroeconomic variables: (1) real return, (2) inflation (both realized and expected),

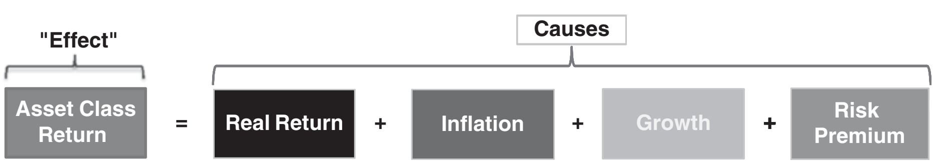


EXHIBIT 6.6 Macroeconomic Drivers of Asset Class Returns

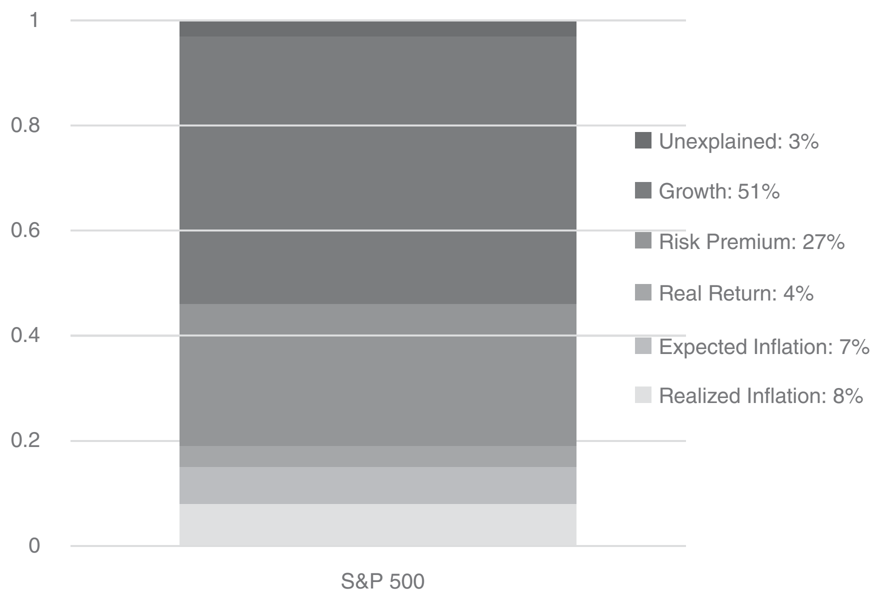


EXHIBIT 6.7 Macroeconomic Drivers of U.S. Equity Returns

(3) growth, and (4) risk premium. These are the global factors that impact the returns to all asset classes.

Not surprisingly, bonds tend to be affected mostly by the real return and inflation rate, whereas more risky assets like public equities are impacted more by growth and risk premium. Exhibit 6.7 shows how the U.S. stock market maps out across these variables. Not surprisingly, public equities map out to mostly growth and risk premium; 78% of the return to the U.S. equity markets can be explained by these two macroeconomic variables. Real rates and inflation also have an impact on the returns to public equities, but their inputs are smaller. Overall, these economic variables explain 97% of the return variability of the U.S. stock market.

Turning to a UHNWI, we include Exhibit 6.8. This portfolio is consistent with a family office whose benchmark is "Don't lose money." One would expect such a goal to have a more conservative portfolio—one tied less to growth assets or risk premiums.

Nearly 60% of this family office is dedicated to real interest rates and inflation. Clearly this is a portfolio that is tied much more to bonds than to stocks or other growth assets. The goal of this family office is not to lose money. It is not surprising that its portfolio would be most exposed to real interest rates and inflation—the key macroeconomic variables that affect bond portfolios.

There is a growing trend of family offices looking through their asset allocation and beta drivers to get a better understanding of what macro factors can impact their portfolio. The reason for this, compared to an endowment, is that family offices often have multiple goals: maintaining a lifestyle, growing wealth, and donating cash and other assets to charity. Given these sometimes competing goals, it is helpful for a family office to look at its beta portfolio to get a better understanding of how the portfolio will react to a spike in inflation, a global economic recession, a general hike in real returns, and so on. With this information, family offices can better adjust their

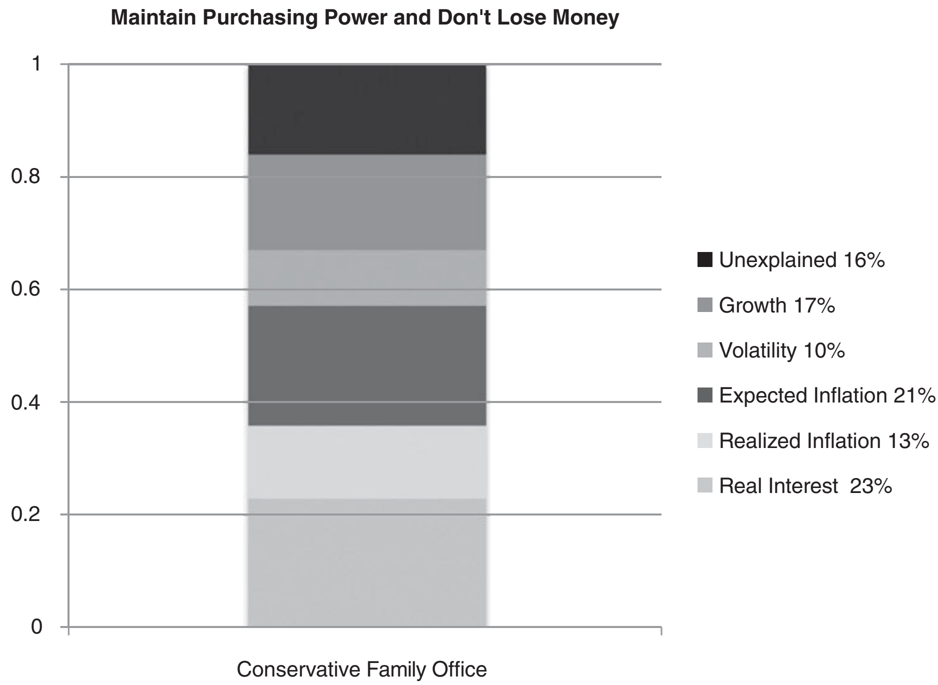


EXHIBIT 6.8 Macroeconomic Drivers of Conservative Family Office Returns

beta drivers to ensure they have the right mix of macroeconomic factors to achieve their individual goals.

6.5 INCOME TAXES OF FAMILY OFFICES

Tax-exempt status is a luxury for endowments, foundations, and pension funds. They enjoy an income-tax-free environment where they do not have to worry about the taxable nature of the gains, losses, and cash yield that flows from an investment. To them a cash dividend is equivalent in tax status to a capital distribution from a private equity investment.

Unfortunately, for family offices, income taxes are a significant constraint. To a family office, there is a distinct difference between cash dividends and capital distributions from a private equity fund. In most jurisdictions, the former are considered ordinary income and taxed at a higher tax rate than distributions from a private equity fund.

6.5.1 Tax Efficiency and Wealth Management

Consider Exhibit 6.9. This shows the payoff profiles for three financial instruments: a swap, a privately negotiated forward contract, and an exchange-traded futures contract. Each of these instruments offers the six-month return on the EURO STOXX Index. If the EURO STOXX Index increases in value over the next six months, each of these instruments will earn a positive return. Conversely, if the U.S. stock market

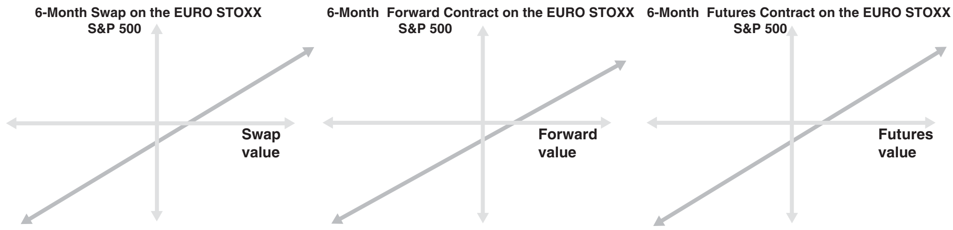


EXHIBIT 6.9 Payoff Diagrams for Swaps, Forwards, and Futures Contracts

declines in value, each of these instruments will have a negative return. In fact, the swap, the forward, and the futures contract all move in lockstep with the up and down movement of the EURO STOXX equity index.

In theory, the payoff profile should show the arrow passing through the zero intercept. However, each of these instruments intersects the *y*-axis slightly below zero. The reason is that there is a cost to each instrument. For example, the swap contract requires the EURO STOXX Index receiver to pay an interest rate, typically LIBOR + X%. Similarly, with respect to the forward contract, the negotiating broker will require the EURO STOXX Index receiver to make a deposit with the broker. Last, for the futures contract, the EURO STOXX Index receiver is required to post a cash amount called the initial margin with the futures broker to ensure that she will honor the contract at its completion.

The main point is that all of these contracts are economically equivalent. An investor should be indifferent between the swap, the forward, and the futures contract. This is true for the endowment, foundation, or pension plan that does not have to worry about the tax consequences of its investments. However, for the family office, there is a very large difference. Taxable investors are concerned with **tax efficiency**, which is the efficacy with which wealth is managed so as to maximize after-tax risk-adjusted return. To pursue tax efficiency, investments should be structured in a way to incur the lowest possible tax liability for a given set of assets and the level of expected return.

6.5.2 Taxability of Short-Term and Long-Term Capital Gains

Forward contracts or six-month swap agreements are considered **short-term capital gains** in the United States, which are trading profits recognized on an investment held for less than one year. Short-term capital gains are typically taxed at ordinary income tax rates, the highest rate currently at 39.6%—effectively 40%—in the United States (not including other potential state and local taxes). The tax rate on short-term capital gains is substantially higher than that of qualified dividends or long-term capital gains in the United States.

Let's assume that the swap or forward contract increases in value by $100,000 over the six-month investment period. Therefore, the net after-tax return to the high-net-worth investor is:

$$\$100{,}000 \times (1 - \text{Tax Rate}) = \$100{,}000 \times (1 - 40\%) = \$60{,}000$$

However, the treatment is different for the futures contract. For example, any exchange-traded futures contract is treated as having 60% long-term capital gains and 40% short-term capital gains, regardless of its holding period. This tax simplification creates an opportunity for sophisticated wealth managers to improve after-tax risk-adjusted returns in the United States and is representative of opportunities that exist in other jurisdictions.

Securities that qualify for this special income tax treatment are known as Section 1256 contracts, and as such, trading in these securities in the United States can be used as a major advantage for high-net-worth investors because 60% of the gains are favorably taxed at the long-term capital gain tax rate, typically about 20% (i.e., half the ordinary income tax rate). In non–Section 1256 securities, the lower tax rate on **long-term capital gains** is levied only on investment gains in which assets are held for longer than one year.

Denoting T_{Ordinary} as the tax rate on short-term capital gains and T_{LTCG} as the tax rate on long-term capital gains, Equation 6.1 provides the blended tax rate on Section 1256 contract gains denoted as $T_{\text{1256 Contracts}}$:

$$T_{\text{1256 Contracts}} = (0.40\ T_{\text{Ordinary}}) + (0.60\ T_{\text{LTCG}}) \tag{6.1}$$

APPLICATION 6.5.2

Suppose that the portfolio of a family office generates \$100,000 of profit from short-term trading and is in an ordinary income tax bracket of 40%. Assume that the long-term capital gains tax rate is half the ordinary tax rate. Compare the after-tax profits if the gains were and were not from Section 1256 contracts. The after-tax profits for non–Section 1256 contracts were shown before to be \$60,000. For Section 1256 contracts, the after-tax profit can be derived by first computing the tax rate using Equation 6.1:

$$\mathrm{T}_{\text{1256 Contracts}} = (0.40 \times 0.40) + (0.60 \times 0.40/2) = 0.28$$

The after-tax gain using the blended rate (0.28) raises the after-tax profits from \$60,000 to \$72,000:

$$\$100{,}000 \times (1 - 28\%) = \$72{,}000$$

Note that 60% of the income is taxed at an assumed long-term tax rate of 20% (half the assumed 40% ordinary rate) and 40% is taxed at the assumed ordinary rate of 40%. The net result is to lower the effective tax rate from 40% to 28%. Clearly, there is an advantage to trading futures contracts and other Section 1256 securities for the family office. This has significant implications with respect to the asset allocation, asset class strategy, and long-term funding goals of the family office. While tax structures vary worldwide, this example clearly shows that advisers, consultants, and family office staff need to understand the implications of different investment vehicles and how each investment affects the specific tax situations of each individual client.

6.5.3 Tax Efficiency and Hedge Fund Investment Strategies

Consider the potential role of hedge funds in improving the income tax efficiency of a family office. Many of these funds use short-term trading strategies such as those associated with relative value and active macro managers who tend to use swaps and forward contracts more frequently than futures contracts. There are three major ways a family office can increase tax efficiency with hedge funds:

1. Focus more on strategies that have a long-term (12-month) horizon to them. This would be equity long/short and event driven.
2. Invest with hedge fund managers who trade more in Section 1256 securities such as futures contracts.
3. Select tax-efficient managers by checking the trading and tax records of hedge fund managers to determine how much of their total return is short-term capital gain and how much is long-term gain.

In the United States, hedge fund managers will provide investors with an annual reporting of the taxable gains and losses for the year. The U.S. K-1 tax form lists the short-term capital gains, the long-term capital gains, return of capital, dividends, and ordinary income receipts from an investment fund over the course of the year. The K-1 form is not particularly relevant for tax-exempt investors such as endowments, foundations, or pension funds. However, for a family office, the K-1 form is an important source of information and a key part of the due diligence process for that hedge fund manager. Family offices routinely check the K-1 forms of prospective hedge fund managers to determine what portion of the hedge fund manager's gains are subject to ordinary income tax rates of 40% and long-term capital gains tax rates of 20%. Some taxable investors may choose not to invest in hedge funds that are likely to issue a K-1 form, as these forms can significantly complicate tax filings, especially when the K-1 forms are issued several months after the typical deadline for filing tax forms.

6.6 LIFESTYLE ASSETS OF FAMILY OFFICES

Another key difference between family offices and university endowments is the allocation to lifestyle assets. **Lifestyle assets or passion assets** include art, homes, wine, airplanes, cars, and boats, where the purchase and collection follows from the lifestyle preferences or the passions of one or more family members. Lifestyle assets can become a significant part of an UHNWI's overall wealth.

6.6.1 Art as a Lifestyle Asset

Art is the best example of a passion or lifestyle asset. Art certainly has been a profitable investment through 2015. In that year alone, one painting by Gauguin sold for almost $300 million (Crow 2015), while a painting by Picasso sold for $179 million (NPR 2015).[7]

There is significant debate about how these assets should be treated in an asset allocation model. Some family offices believe that art, for example, is an asset class that should be included in the asset allocation process. The reasoning is that art is a scarce commodity and typically increases in value over time. In fact, large money center banks such as Citicorp and JPMorgan even provide an outsourced curator function to look after a family's artwork.

More to the point, there are art indices that attempt to track the return to art. One index uses repeat sales of the same painting or sculpture. By using transaction pairs of the same artwork, the issue of heterogeneity or uniqueness is removed as a bias in the index calculation.[8] One problem with such a pairs index is that the time interval between the "paired trade" can be many years—even decades. Furthermore, museums are active purchasers of artwork, and these purchases may permanently remove the art from the trading market. Also, this type index does not take into account the vast majority of single transactions in the art world. Last, this type of index ignores private transactions that account for up to 50% of all art transactions.[9]

Another type of index attempts to measure the premium paid for auctioned artworks above the estimated auction price quoted by dealers and auction houses. This index measures a rarity premium as the difference between the realized price at auction and the estimated value as established by the auction house. While this type of index attempts to be more empirical, ultimately, the rarity premium is based on the subjective estimate by a dealer or auction house of the estimated sales price for the artwork.[10]

Other family office managers take the view that art is not an asset class because there is no underlying systematic risk premium that can be used to risk budget art as an asset class. Without an underlying risk premium, it is hard to determine how art can be built into an efficient frontier of portfolio construction. Furthermore, unlike the homogeneous assets such as stocks and bonds, each artwork is unique and is defined by a set of physical and intangible characteristics.

Another argument against art as an asset class is that there is no cash flow associated with a piece of art. Therefore, a discounted cash flow model cannot be used to determine its present value. Similarly, it is very difficult to measure the underlying volatility of artwork—individual pieces trade infrequently such that there are very few data points upon which to build a distributional analysis of the returns to art. Last, art is an illiquid asset class; most family offices do not trade in and out of art. They purchase art for the personal enjoyment of the family without regard to trying to generate a capital gain.

6.6.2 Lifestyle Wealth Storage and Other Costs

Financial assets typically have little or no storage costs (custody costs) and in some cases net benefits from securities lending. Expense management can be an important component of lifestyle asset management. While lifestyle assets typically do not generate cash flow, there are ongoing costs of ownership. Expenses such as insurance, storage, curation, security, and general maintenance can increase rapidly as a collection of passion assets grows. And it is not just guarding against theft; for older and rarer collectibles like paintings or wine, the assets must be guarded against the elements like sunlight, dust, and water damage. The proper installation of the passion asset can also be expensive.

In order to reduce these costs and potentially generate cash flow, some art investors have chosen to lease their art, as described by Harris (2015). Demand for leasing can come from corporate art collections, which gives the advantage of tax-deductible lease payments. Other lessees may include art investors who may wish to display art that they cannot afford to purchase or may wish to live with a piece of art before making a final purchase decision. Perhaps, leasing art could lead to greater appreciation, such as when a work is displayed in a museum that leads to a larger number of viewers, some of whom may potentially be interested in purchasing the piece. While leasing models vary, some lessors may intend to sell the art and may credit all or part of the lessee's payments toward the eventual purchase of the piece.

A significant trend is that artwork is often held in "free ports." **Free ports** are specialized, climate-controlled repositories for art and other valuable goods belonging to the very wealthy—similar to a custody bank for stock certificates. Not surprisingly, the best-known free ports are in Geneva and Zurich, Switzerland, home to many of the oldest and most private of wealth management firms. Additional free ports exist in Beijing, Singapore, and Luxembourg. More and more, UHNWIs may see art as a portable store of value where the physical display of the artwork is not necessary to achieve personal enjoyment or social status.

Some free ports, such as Singapore, offer complete confidentiality with regard to the nature of the art, its value, and the identity of the owner. However, other free ports, such as those in Zurich and Geneva, are governed by laws that require descriptions, values, and country of origin for the art stored.

Free ports have an important advantage for art collectors. With UHNWIs and museums, foundations, auction houses, and other private collectors, free ports have become an acceptable place of business for the purchase, sale, and transfer of art ownership. The key advantage is that free ports charge no sales taxes—taxes are charged by the country of destination when the art leaves the free port. Until then, while stored at the free port, the goods are considered "in transit" and typically not taxed. This makes free ports a convenient and often a permanent place of storage for art.[11]

6.6.3 Lifestyle Assets and Portfolio Management

Following this discussion, one way that lifestyle assets are used in portfolio construction for a family office is as a constraint on the asset allocation process. Exhibit 6.10 demonstrates this constraint. It shows three broad buckets of assets that are used to achieve three family member goals: (1) maintaining a lifestyle, (2) wealth growth, and (3) wealth transfer.

For this individual two different portfolios are shown: a moderate portfolio and an aggressive portfolio. In each case, the lifestyle assets remain constant. It is the other two buckets of assets, wealth generation and wealth transfer, that must adjust to achieve the moderate and aggressive portfolio allocation.

Sometimes these two buckets are referred to as **balancing portfolios**, which are used as counterweights to a pool of assets that, for some reason, cannot be adjusted itself. In this case, the lifestyle assets are never expected to be sold or rebalanced. They exist for the personal enjoyment of the family member. Consequently, the wealth generation and wealth transfer buckets are used to balance the overall risk profile for the family member.

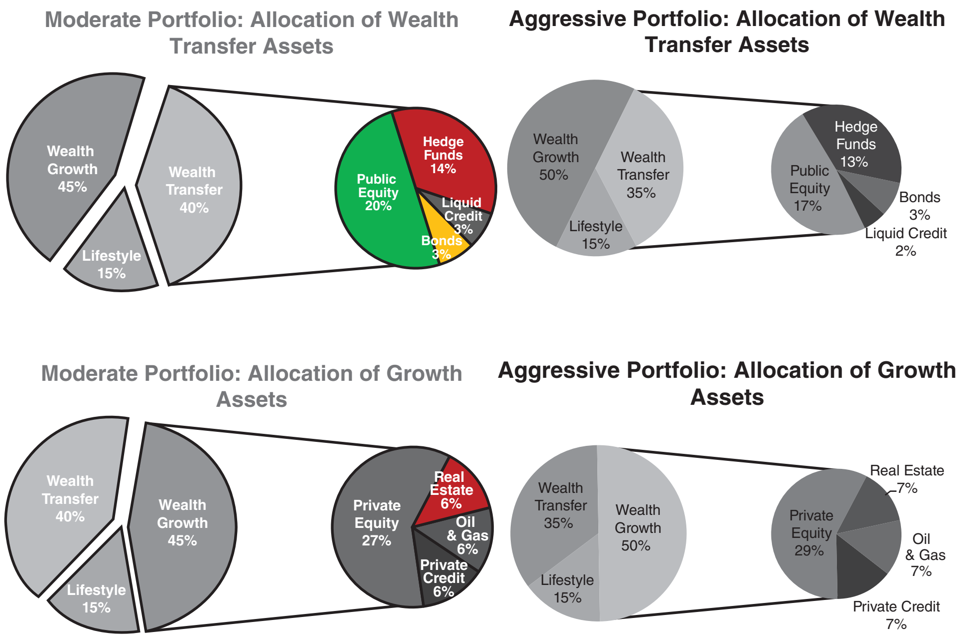


EXHIBIT 6.10 Asset Allocation Including Lifestyle Assets

6.6.4 Concierge Services

Closely related to lifestyle assets is a growing part of family office management known as **concierge services**, where the family office will attend to mundane details that most people have to deal with in their daily lives such as personal shopping and travel arrangements. While concierge services may be offered by family offices, the ultra-high-net-worth individuals can afford to pay a staff directly to manage these details for them rather than receiving them in conjunction with their wealth management services. A sample list of concierge services (but by no means all-inclusive) follows:

- Personal shopping—everything from daily groceries to wine purchases to private showings of current fashions in salons and boutiques
- Art curation and purchase, including attending art auctions as well as private donations to museums
- Travel arrangements, which can include personal guides for climbing Mount Everest to planning a trip to Disney World with the children or grandchildren
- Entertainment and sporting events—front-row seating, behind-the-scenes private tours, meeting with the rock stars and athletes
- Purchase, sale, and maintenance of cars, both modern and antique
- Jet and yacht rentals, purchases, and maintenance, including hiring staff to fly the planes or sail the yachts as well as storage costs for airplane hangars and docking of boats
- Setting up concierge medical services. This is a fast-growing part of family office services—arranging medical visits with first-class physicians without the hassle of medical insurance roadblocks or waiting in doctors' anterooms.

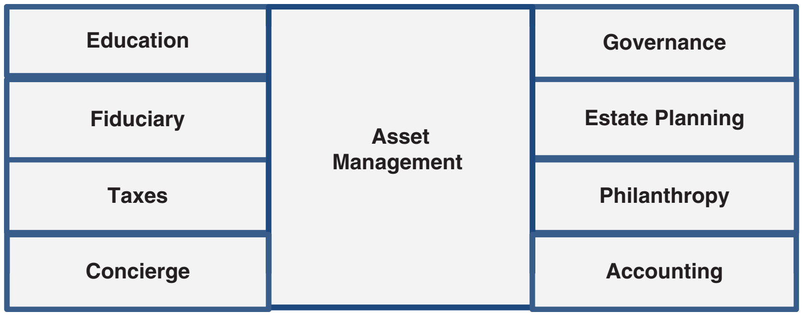


EXHIBIT 6.11 Services Offered by a Family Office

Another part of concierge services is directly related to the wealth management. These include tax preparation, administration of the assets, custody, estate planning, education, and philanthropy. Exhibit 6.11 shows the complete range of a full-service family office.

6.7 FAMILY OFFICE GOVERNANCE

Family offices operate very differently from one another. Again, there is no endowment model counterpart to family office management. In the endowment model, independent trustees form the bulk of the endowment board with, typically, only one university employee sitting on the board.

6.7.1 Governance Structures of Family Offices

Some family offices have a formal structure that includes a board of directors or trustees, much like an endowment or a foundation. However, in the family office case, almost all of the board members are family members. There may be one or two independent board members, but this is limited because of the confidentiality of the family wealth. In this case, there are formal board meetings, usually quarterly, where investment memos are proposed and vetted by the board of family members.

Other family offices have a less formal structure and are often controlled by a single person. This is especially true when the first generation that created the family wealth is still alive, has sold the family operating company, and now devotes his or her time to the family wealth. Usually, this is a patriarch or matriarch who forms the core of decision-making and governance for the family. In this case the investment committee is typically the patriarch/matriarch and the chief investment officer.

Like everything else in life, there are pros and cons to formality in family office governance. The more formal style ensures a rigorous process with detailed investment memos to support every opportunity. Also, there is the forum for debate as well as ensuring that every family member's investments and interests are attended to fairly and completely.

On the other hand, the patriarch/matriarch model offers more flexibility. Investment committee meetings are whenever the patriarch/matriarch wants them or needs them. Decision-making is quick and decisive, and flexibility is an advantage.

There is no single model that is right for all situations. The more formal model is typically set up by the time the wealth is passed to the second generation unless a member of the second generation is anointed to be the new patriarch/matriarch of the family. It really depends on the individual family's comfort zone with a patriarch/matriarch model versus a board of trustees model and the natural equilibrium for managing the family office depending upon the generation cycle of the family itself.

6.7.2 The Challenges of Family Wealth Sustainability

A common phrase in wealth management is "shirtsleeves to shirtsleeves in three generations," which acknowledges the fact that the vast majority of family fortunes never last beyond the third generation after the wealth was earned, as the first generation earns the wealth, the second preserves it, and the third spends it or distributes it across a large number of descendants. **Dynastic wealth** is an amount of wealth so large that is has substantial potential to be maintained for a large number of generations. The difficulty of maintaining dynastic wealth is evidenced by the fact that none of the 20 richest families in the United States in 1918, such as Ford, Rockefeller, or Carnegie, were among the 100 richest families in 2012 as measured by *Forbes*.[12]

There are four primary factors related to this loss of family wealth over time:

1. A growing number of descendants as the number of generations increases, which dilutes wealth among a larger number of people
2. A lack of skill or interest in the family business by members of the younger generations
3. A lack of education and planning with how to prepare younger generations to be productive and understand how to handle their share of the family wealth
4. An intentional focus on philanthropy, where the assets are distributed to benefit broader society rather than being left for descendants

These types of issues are described in detail in Daniell and McCullough (2013), who believe that it may be easier to become wealthy today than to stay wealthy over multiple generations.

6.7.3 Strategies to Maintain Family Wealth

In order to address the issue of diluting wealth across generations, it is recommended that wealth be held in a multigenerational trust controlled by the family office or the family business. Similar to an endowment, the assets are professionally managed with a focus on total return, while income to the descendants is limited to a sustainable spending rate that will allow the assets to maintain their value over multiple generations. Limiting overspending by family members and explaining a sustainable distribution rate are important factors in maintaining the wealth over time. The income or assets distributed per beneficiary inevitably decline over time, as there could be 60 beneficiaries, or descendants who receive a share of the family wealth, over the course of five generations or 120 years. With an annual spending rate of 2% to 4% and adequate investment returns, the family office may be able to operate in perpetuity, distributing income to many generations of descendants. The assets will be depleted quickly, perhaps within two generations, if the annual spending rate is 6% to 10%, which likely exceeds the real return to the portfolio.

Ideally, the trust would be protected from issues such as family disputes and divorces (dissolutions of marriage) that are well known to quickly reduce family wealth. A properly structured governance process would educate family members on the issues of wealth management and sponsor regular discussions designed to reduce the probability of conflict within the family over time. If the family members agree on the vision, values, and goals of the family, the managers of the family office can structure operations in a way designed to meet these goals.

Managers of the family office may also be responsible for governance issues, planning family meetings, and educating family members on investments and wealth management issues. Records within the family office can preserve and document the family vision, culture, and wishes over time. This institutional memory is very important when the family office changes financial advisers, accountants, and attorneys.

6.7.4 Family Office Inheritance and Succession Strategies

One of the key questions is determining a strategy for **inheritance**, which is the distribution of assets after the death of members of the older generation. Loomis (2012) quotes Warren Buffett as saying that children should inherit "enough money so that they would feel that they could do anything, but not so much that they could do nothing." Warren Buffett is planning to donate nearly all of his wealth to causes outside of his family, leaving his family with millions and his philanthropic causes with billions. Young family members should be educated on the ways of wealth on a continuous basis over time in a way that they will be prepared for when they receive a distribution of wealth. They receive education not only on wealth management, but also on ways of exercising diligence, discretion, and the avoidance of destructive behaviors. Realistic expectations on the size of the inheritance can serve to reduce anxiety and division within the family. Daniell and McCullough (2013) explain that giving too much wealth too soon may reduce the work ethic of the beneficiary and create a sense of entitlement. To overcome these issues, it is recommended that wealth be distributed over time, ideally after the beneficiary reaches a series of milestones. Rather than giving $10 million to children once they turn 18 years old, perhaps $1 million is distributed at age 25 or 30 after they have graduated from college and worked for three to 10 years in a non-family business proving that they can balance a budget on their own income. Further distributions, perhaps $10 million over the course of a lifetime, may take place as the beneficiaries continue to participate in society in a productive way, often at milestones such as when they are married, have children, buy a house, or need to fund their children's education. Some families choose to distribute assets to beneficiaries in a way that matches their income, as those with higher levels of income receive a larger share of the family wealth, as they have proven to be responsible and productive in their own affairs. It is important that all descendants understand their role in the stewardship of the family fortune.

Some believe that those who materially participate in the family business and in the governance of the family office should receive a substantially larger distribution than those who choose not to contribute in this way. This participation in the family business and family office would commence after the completion of their education as well as professional experience in the non-family business. Once the children have proven themselves both externally as well as in the management of the family business and family office, the **succession planning** process starts, which is the process of naming a new leader of the family business and potentially a new governance structure after the death or retirement of the founder or current leader of the business. Succession planning is important, as Daniell and McCullough (2013) state that 92% of family businesses do not last three generations and there is less than a 30% probability of maintaining family wealth across three generations. Training outside the family business is vital, as external experience can bring a greater diversity of views and experience into the family business once the family members complete their external training and move into the family business.

6.8 CHARITY, PHILANTHROPY, AND IMPACT INVESTING

When members of the oldest generation die, their assets are divided between tax liabilities, inheritances distributed to family members or other persons, and donations to benefit society beyond the family.

6.8.1 Charity and Philanthropy

Family estate planning is the process for planning the distribution of assets upon the death of preceding generations. The estate plan considers the goals of the family, the number and type of beneficiaries, the taxable nature of wealth upon death, as well as insurance and giving strategies. Distributing assets to family was discussed in the previous section, while tax and insurance strategies may vary widely across families and domiciles. **Estate taxes** are levied in many jurisdictions by governments on accumulated wealth after the death of its owner. While assets distributed to family members are often subject to these taxes, the estate tax law in many jurisdictions allows the distribution of assets outside of the family on a pre-tax basis. This allows families to maximize their donations to institutions that seek to benefit society.

Many families may choose to distribute all or part of their fortune to others outside of the family, using their wealth to benefit society. There are a variety of causes to which families contribute, but they tend to be focused on religion, health care, education, arts and museums, poverty and community development, as well as the environment and issues of social justice or inequality. Each family may have a different area of passion, and the family office team can assist the family in defining how and where this wealth can be distributed to best benefit the selected cause.

There are two terms for donating assets to benefit society: charity and philanthropy. Frumkin (2006) believes that these are very different concepts, even though they may be used interchangeably in conversation. **Charity** is the giving of money or time to social causes, typically to meet more immediate needs and without accountability on behalf of the recipient. For example, a donor may make a charitable gift to a homeless shelter that feeds and houses those with immediate need, while the shelter does not seek to overcome the situation of poverty through counseling or job training. Without addressing the root causes of poverty, the poverty is sustained and the recipients of the charitable gifts may be in need for long periods of time. In contrast, **philanthropy** is the giving of money or time with the intent of making a lasting change. In the example of the homeless shelter, a philanthropic gift would seek longer-term benefits to permanently improve the situation of the homeless by providing the counseling, education, and job training needed to reintegrate the homeless back into productive society. Philanthropy is designed to help people to help themselves and demand accountability and effort as to the ultimate impact of the gifts.

Frumkin (2006) discusses philanthropy in the context of Andrew Carnegie, who stated that "the man who dies rich, dies disgraced." Carnegie believed that wealth should be distributed during the life of the donor, as the donor has no need for the assets after death. Carnegie is joined by 21st-century philanthropists Bill Gates, Warren Buffett, and Mark Zuckerberg, who are seeking to improve outcomes in global health and education during their lifetimes. Buffett has also encouraged other billionaires to leave nearly all of their wealth to philanthropic causes. Azim Premji, the chairman of Wipro Limited, donated a significant portion of his wealth to improve education in India. When signing Warren Buffett's giving pledge, he stated, "I strongly believe that those who are privileged to have wealth should contribute significantly to try to create a better world for the millions who are far less privileged."[13]

These men have amassed great wealth during their lifetime, but they have also developed global influence due to their business acumen. Rather than giving only money, many successful businesspeople seek to use their social and political influence and business skills to measure and improve the impact of their philanthropic ventures. Andrew Carnegie, along with John Rockefeller, had an interest in scientific philanthropy, which searched for the underlying causes of poverty and donated to organizations that were able to address these causes. If successful, the need for charity and social welfare would decrease. Ideally, philanthropy would allow recipients to become gainfully employed, ultimately becoming self-supporting and eventually accumulating their own wealth. Some philanthropists sponsor public policy research institutes that actively research ways in which to better the world. Through publication of the research and conversations with lawmakers, the institutes seek to influence legislations and regulations that would further their social agenda. For example, funding scientific research into climate change could influence government officials to create and enforce regulations that could reduce pollution and improve the environment. Many problems are so large that they require changes in governmental regulations in addition to substantial investments of individuals as well as governmental entities. Just as venture capitalists may seek to invest in disruptive technologies to meet business goals, philanthropists are seeking to promote innovative thinking into solving global issues.

6.8.2 Impact Investing

The World Economic Forum (2014) defines **impact investing** as an investment approach that seeks to earn financial returns while generating measurable and positive social impact. While philanthropy and charity donate assets to worthy causes, impact investing seeks to earn a positive financial return and a return of capital while using the investment to drive social benefits. Impact investing is closely related to concepts such as socially responsible investing (SRI) and environmental, social, and governance (ESG) investing. The World Economic Forum estimates that 17% of family office assets are now allocated to impact investing, with interest greatest among women and younger family members who desire to use their wealth to make a difference in the world. In 2013, the *Financial Times* estimated that 68% of family offices have some involvement in impact investing.

Historically, many investors were introduced to this topic by using their social goals to tilt investments in their public equity portfolio. In **negative screening**, investors intentionally eliminate companies from their portfolio that are deemed to have a negative impact on the world, such as from pollution, harmful products, or unfair employment practices. A newer concept is **positive screening**, where companies are added to the portfolio when they are perceived to do good in the world, such as producing helpful products and having high wages and benefits and good working conditions for their workers and suppliers worldwide, even in areas of extreme poverty. The results of either version of screening have not been extremely beneficial, as the return and risk of screened portfolios may not provide stronger financial results than an indexed portfolio, and it is unclear how changes in portfolios of public companies impact the behavior of these very large and highly diversified firms.

As a result, many investors believe that the social impact can be magnified through focused investments in private equity and private debt structures. There is a wide variety of risk and return targets from private impact investments. Some investments are **impact first**, where investors have a greater focus on the social good of their investments, and may accept projects with higher financial risk or lower financial returns. In an extreme case, investors may buy bonds with a zero yield and a simple expectation of the return of capital if the proceeds of the debt offering are used to meet their social goals, perhaps building affordable or environmentally efficient housing. **Finance first** investors would like to earn an investment return competitive with market returns and commensurate with the risk of the investment and place relatively less priority on social impact. Once investments with competitive risk-adjusted returns are identified, finance first investors select the projects that most benefit the cause they have chosen to support. Some investors espouse impact alpha. **Impact alpha** is the theory that ventures choosing to do the right things from a social perspective will ultimately be rewarded in the marketplace with above-market financial returns, or that ventures that have socially objectionable operations have substantial risks of generating below-market financial returns. For example, a firm with operations that are negative for the environment may be faced with substantial future cleanup costs that are not currently factored into the stock price.

Impact investing has long had a reputation for earning below-market financial returns. If this is true, the assets allocated to these strategies by fiduciaries such as pension plans will be limited. However, if it can be proven that impact investments provide competitive risk-adjusted returns, then the amount of assets dedicated to impact investments may increase substantially. This increase in assets may start first with family offices and then move into endowments and foundations, many of which are already affiliated with organizations charged with benefiting the social good. Pension plans and others with the largest pools of capital will likely follow.

The Case Foundation (2015) notes that impact investments can be made across asset classes, many of which are clearly alternative. These include public and private equity, fixed income, real estate, and infrastructure. Investment structures can include funds, direct investments, or structured products such as community development bonds.

A key goal of impact investing is to be able to measure the outcomes of these investments. Measurement can be easier when the cause is well defined and the investment is relatively small and where the investor can have an influence on operations, if desired. Some examples of ways of measuring the impact of social investments are the number of jobs or housing units created, the decline in pollution or increase in the availability of clean water, or the number of children vaccinated against diseases. To the extent that investors become involved in the cause, their business acumen and political influence may further enhance the productivity, measurability, and long-term impact of these investments.

6.9 TEN COMPETITIVE ADVANTAGES OF FAMILY OFFICES

Almost every investor has some form of natural advantage that can be exploited as part of the portfolio management process. Pension funds, for example, have huge balance sheets that can bear a lot of liquidity risk over a long period of time. Endowments and foundations are tax-exempt so they can seek out hedge fund strategies without regard to whether short-term or long-term capital gains are generated.

Similarly, family offices also have some natural advantages that can help manage the overall portfolio. For purposes of this discussion we will refer to second- and later-generation family offices—after the initial wealth has been generated and the family operating company has been sold.

1. Aggressive asset allocation. In a prior chapter it was noted that one of the reasons for endowment outperformance compared to other investors was an aggressive asset allocation geared toward private equity, venture capital, hedge funds, and other alternative investments not necessarily available to average investors. However, as indicated in Exhibit 6.4, UHNWIs can have asset allocations that are even more aggressive than university endowments. This flows from their ability to gain access to the best managers as well as their larger risk appetite.
2. Liquidity premium capture. Exhibit 6.5 also showed that the liquidity profile for a UHNWI can be more aggressive than a well-known university endowment. This aggressiveness allows UHNWIs to capture the liquidity premium associated with illiquid asset classes. UHNWIs are in an enviable position because they don't have a university budget to backfill. Endowments are constrained by the amount of the university budget to which they must contribute each year. Family offices do not have this constraint, and consequently they can commit even more of their wealth to long-dated assets compared to an endowment.
3. Deal flow. In addition to allocating to private equity and venture capital funds, family offices and UHNWIs are often approached by entrepreneurs and operating companies for direct infusions of capital. Most family offices prefer to make direct private investments in existing, operating companies with established cash flows rather than start-up ventures, but there is no hard-and-fast rule. Direct deals flow from the family's own network of contacts in the banking industry, other private family offices, and lawyers, accountants, and other service providers. Some family offices establish a reputation for direct deals that attracts even more operating companies seeking private capital.

   From an operating company's perspective, billionaire capital is an excellent source of funding. This capital tends to be long-term and the family office is not looking to change the management of the company or the board of directors—unlike private equity managers, who often cause significant turnover and turmoil within an operating company. Also, the family office can often make introductions to other investors or customers for the company.

From the family office perspective, with a direct deal, there are no "2 & 20" fees to be paid to a private equity manager. In addition, the family office gets complete transparency into the underlying investment, unlike allocating capital to a private equity fund where the family office has no control over the investments that will be made by the private equity manager. Last, the family office often contributes the private capital at a discount to the company's enterprise value, unlike private equity managers who typically have to pay a premium to acquire the company.

4. Speed. Family offices can act quickly to jump on opportunities that a pension fund or an endowment would have to pass on because it cannot process the investment quickly enough. Pension fund and endowment governance requires committee approval to make investments. This can take several months to schedule the investment opportunity on the investment committee agenda. By contrast, a family office can quickly convene family members to review an opportunity, typically in a matter of days rather than months.
5. Governance and management of assets. Managing individual wealth can be complex because it needs to be transparent and devoid of conflicts of interest. An internal family office solves these problems by addressing the individual goals of the family members: providing tailored advice within a framework of complete confidentiality. In addition, the family office can blend the goals of the individual family members with the overall objectives of the family office: cost savings, central control of the family wealth, and sophisticated asset management.
6. Alignment of interests. When a wealthy family allocates capital to different financial advisers and managers, it will rarely lead to a perfect alignment of family interests. In fact, it often leads to competition among the outside money managers and behavior that can be counterproductive. A family office eliminates these potential conflicts by creating structures where the interests of the family members, the family office, and external money managers work together toward common goals. For example, by pooling the money of the total family, the family office can often negotiate fee breaks and other advantageous terms that a single family member may not be able to access. This has also led to family offices bringing more asset management in-house to avoid external conflicts of interest as well as harder negotiation of fees and promotes.
7. Higher returns. The ability to manage the assets of the family in a centralized and professional office can generate the benefit of achieving higher returns. As previously mentioned, the family office can often negotiate fee breaks that a single family member cannot. Every basis point saved in fee negotiations goes straight to the bottom line of additional returns. Also, a central management structure for the family's wealth allows for economies of scale that generate a more efficient and cost-effective management of the family wealth. Last, internal professional management, devoid of any conflicts of interest, gives the family the best chance of achieving higher returns.
8. Risk management. Centralization of asset management also has the benefit of centralization of risk management. All of the same benefits flow from this:

cost-effective management, better alignment of interests, and ability to manage risks across asset classes and time horizons. Another key benefit is the centralization of risk reporting and performance attribution—a more cost-effective way for the family to ensure that their wealth is managed intelligently and prudently. The financial crisis of 2008 taught the need for intelligent risk management and reporting. Family offices are no different from other investors in this regard, and this has become a key component of an integrated family office. Since the 2008 crisis, family offices are putting more emphasis on risk management, as evidenced by more diversified investment portfolios and the use of hedging tools—particularly as it relates to a concentrated asset like a family operating company.

9. Centralization of services. The family office can also centralize all of the other services demanded by a UHNWI: estate planning, taxes, accounting, travel, art curation, philanthropy, and so on. The centralization provides two key benefits: cost savings and confidentiality. While it has been mentioned several times, it is worth emphasizing that the overwhelming majority of family offices prefer to remain quiet, confidential, and out of the press. It is only the most flamboyant personalities that seek to publicize their investments, philanthropy, and other lifestyle choices. For the most part, billionaires are quiet, even shy people who want to have their wealth managed discreetly, and their philanthropy to be offered anonymously.
10. Lifestyle assets. Lifestyle assets flow from some passion of the family member. Art, wine, and other collectibles are prerogatives of the very wealthy. However, there is no consensus about how these assets should be accounted for as part of a well-diversified portfolio. One way to account for these assets is to use them as a constraint in the asset allocation process—a dedicated block of assets that provide a fixed allocation in the portfolio construction process.

6.10 CONCLUSION

This chapter has explored the role of the family office in wealth management, as well as in the alternative investments universe. Previously, we discussed tax-exempt investors such as endowments, foundations, and pension funds. Family offices can have quite different concerns from other large investors, as taxes and the changing composition of family members over generations are key concerns. Family offices not only address issues of wealth management, but also seek to educate family members on matters of wealth and to assist in succession planning for family wealth and the family business.

NOTES

1. Kindle (2013).
2. See Capgemini (2012).
3. See Amit et al. (2011).
4. This survey was conducted by the author as part of his own network of UHNW family offices
5. See Collins (2015)
6. See Amit et al. (2011).

7. See Crow (2015).
8. See the Mei Moses All World Art Index, Mei and Moses (2002).
9. See Geman and Velez (2015).
10. See Geman and Velez (2015).
11. See Geman and Velez (2015).
12. *Forbes* (2002); Kroll (2012).
13. India Today, February 23, 2013. http://indiatoday.intoday.in/story/azim-premji-wipro-chairman-donates-millions/1/251436.html

REFERENCES

Amit, R., H. Liechtenstein, M. J. Prats, T. Millay, and L. Pendleton. 2011. "Single Family Offices: Private Wealth Management in the Family Context." Wharton Global Family Alliance.

Capgemini. 2012. "The Global State of Family Offices." Internal white paper.

Case Foundation. 2015. "A Short Guide to Impact Investing." October.

Collins, M. 2015. "Private Investment Firms Win the Right to Keep Money in the Family." *Bloomberg News*, February 10.

Crow, K. 2015. "Gauguin Painting Sells for the Record Sum of Almost $300 Million." *Wall Street Journal*, February 6.

Daniell, M. H., and T. McCullough. 2013. *Family Wealth Management*. Hoboken, NJ: John Wiley & Sons.

Forbes. 2002. "The First Rich List." *Forbes,* September 27.

Frumkin, P. 2006. *Strategic Giving: The Art and Science of Philanthropy.* Chicago: University of Chicago Press.

Geman, H., and T. Velez. 2015. "On Rarity Premium and Ownership Yield in Art." *Journal of Alternative Investments* (Summer): 8–21.

Harris, G. 2015. "Something Borrowed: The Growing Industry of Art Leasing." *Financial Times*, June 12.

Kindle, N. 2013. "Family Office Market Growth, Size, Share and Forecast Research Report." Culrav.org.

Kroll, L. 2012. "Forbes World's Billionaires 2012." *Forbes*, March 7.

Loomis, C. 2012. "Should You Leave It All to the Children?" *Fortune*, November 21.

Mei, J., and M. Moses. 2002. "Art as an Investment and the Underperformance of Masterpieces." *American Economic Review* 92.

NPR. 2015. "Picasso Painting Sells at Auction for $179 Million, a Record." May 11.

World Economic Forum. 2014. "Impact Investing: A Primer for Family Offices."

PART 2

Private Equity

CHAPTER 7

Private Equity Market Structure

The growing interest in private equity (PE) investing has arisen in part as a result of its potential to earn superior long-term returns when compared to those of public equity investing and in part due to the diversification benefits it provides. Investments in PE funds offer access to privately held companies not available in the traditional investor landscape and to the expertise of intermediaries (PE managers) in creating value by proactively influencing the management and operations of these companies.

Institutional investors typically focus on the **organized PE market**, where professional management is provided by intermediaries. There is also an **informal PE market**, which comprises angel capital and what is often referred to—not without justification—as family, friends, and fools. Companies can also receive funding in the form of the founder's savings and efforts, commonly known as blood or sweat equity. The number of investments made in the informal PE market is probably several times larger than that made in the organized PE market; however, it is difficult for institutional investors to gain the information and access necessary to invest in this informal market effectively.

7.1 MAIN STRATEGIES OF PRIVATE EQUITY INVESTMENT

PE funds refer to a multitude of investment strategies with varying risk-return profiles (see Chapter 11) and liquidity profiles (see Chapter 13). For the purpose of this discussion, we differentiate among the three primary, and most important, types of strategies: venture capital, buyout, and mezzanine. These strategies form the bulk of a typical institutional investor's PE portfolio.

Venture capital (VC) relates to equity co-invested with entrepreneurs to fund their young and potentially fast-growing companies and is often active in technology sectors such as telecommunications, life sciences, and clean technology. Venture capital has two subcategories, depending on the stage of development of the funded company:

1. **Early stage:**[1] This stage is split into seed stage and start-up stage. The seed stage takes place before a company is set up and any new product is sold. The financing provided is used to fund research, to assess an initial concept, and to develop a new product. Once successful, further financing is provided during the **start-up stage** to establish the company and begin to market its new product.

2. **Expansion stage**: A company in this stage (also called the development capital stage), which may or may not have reached profitability, has already established the technology and market for its new product. The financing provided is used to allow greater or more rapid growth by increasing production capacity, developing markets or products, or providing additional working capital.

VC investments are not comparable to traditional financial assets, such as public equity or bonds, and have characteristics that make it difficult to apply traditional portfolio management techniques. These investments are still generally in the cash-burning stage and may be several years away from profitability.

Buyout[2] relates to capital provided as a mix of debt and equity to acquire from current shareholders an established business, business unit, or company (generally privately held or a spin-off from a large private or public company). *Buyout* is a generic term that denotes a change of ownership with the support of PE investors. A management buyout (MBO) occurs when the current management acquires the company, whereas a management buy-in (MBI) takes place when new managers come from outside the company. When a public company is bought entirely and delisted from the stock exchange, the transaction is referred to as public-to-private (P2P). In buyout funds, portfolio companies are established, have tangible assets, and are normally beyond the cash-burning stage, which allows the use of debt to finance part of the transaction. In these cases, buyouts are referred to as leveraged buyouts.

Mezzanine relates to capital provided through the issuance of subordinated debt, with warrants or conversion rights to finance the expansion or transition capital for established companies (usually privately held or below investment grade, or both). Mezzanine financing is halfway between equity and secured debt. While mezzanine financing gives a more predictable cash-flow profile, it is unlikely to provide capital returns comparable to other PE financing forms.

Beyond these three strategies, other private equity strategies exist:

- **Rescue** (or turnaround) refers to a strategy in which capital is provided to help established companies recover profitability after experiencing trading, financial, operational, or other difficulties.
- **Replacement capital** (also called secondary purchase) refers to a strategy in which capital is provided to acquire existing shares in a company from another PE investment organization.

7.2 MAIN DIFFERENCES BETWEEN VENTURE CAPITAL AND BUYOUT

VC and buyout transactions differ in several significant aspects, notably their business model, their deal structuring, the role of the PE manager, and valuation. These and other differences are summarized in Exhibit 7.1 and discussed in the following sections.

The classic argument presented for diversifying among PE classes, and especially between VC and buyout strategies, is that various PE strategies may exhibit low or even negative correlations with one another and differ in terms of growth and value investing. To begin with, buyout transactions are largely debt financed and tend to

EXHIBIT 7.1 Venture Capital–Buyout Comparison

	Buyout	Venture Capital
Sector	Established industry sectors	Focus on cutting-edge technology or rapidly growing sectors
Stage	Stable growth and mature stages	Seed, start-up, and expansion stages
Approach	Financial engineering, corporate restructuring	Industry know-how, product development and commercialization
Uncertainties	Risk is measurable	Risk is difficult to measure (uncertainty)
Source of returns	Leverage, company building, multiple arbitrage	Company (and market) building, finding follow-on investors
Selection	Intensive financial due diligence	Limited financial due diligence but extensive sector/product due diligence
Valuation constraints	Cash-flow projections overlooked by credit lenders	None; often no non-VC third party oversight
Business model	High percentage of success with limited number of write-offs	A few winners with many write-offs
Financing	Club deals and large investment	Limited syndication; several investment rounds
Monitoring	Cash-flow management	Growth management
Success factor	Backing experienced managers	Backing entrepreneurs

perform well during depressed public equity market periods, when debt is cheap. However, if depressed equity prices are accompanied by a widening of credit spreads (e.g., during the financial crisis of 2008–9), then leveraged buyout transactions may not be feasible. Second, venture capital relies on the stock market as the most profitable exit route; therefore, when close to exit, VC often shows strong correlation with small-cap indices. Consequently, venture capital would be expected to do better during equity bull markets when initial public offerings (IPOs) as well as trade-sale activity is more robust. Historically, buyouts have provided more stable returns with an orientation toward minimizing risk, whereas venture capital has occasionally produced higher rates of return in certain markets but brings the possibility of higher losses. Thus, investors seeking long-term stable returns would be inclined to overweight buyouts, whereas those seeking higher returns would do so through increased exposure to venture capital.

7.2.1 Business Models of Venture Capital and Buyout

Attractive VC investment opportunities can be difficult to assess and are usually concentrated in a few high-technology sectors, which often results in a relatively high number of small investments. Returns stem from taking large risks to develop new businesses and concentrating efforts and capital through several incremental funding rounds. The goal is to build companies that can be sold or taken public with a high multiple of invested capital. These few big wins need to compensate for many failures.

VC-funded companies can be seen as works in progress, with intermediate stages of completion. These stages of completion are often distinguished by milestones, such as rounds of financing (rounds A, B, and C) or, in the case of biotech companies, phases of clinical trials (phases I, II, and III). In this respect, they are development projects that cannot be prematurely exited without risking the loss of most, if not all, of one's invested capital. Thus, VC transactions should be viewed as long-term investments.

Large capital requirements and lower risk levels result in most buyout managers making a smaller number of investments compared to venture capitalists. A multitude of approaches can be combined in a transaction, such as divestment of unrelated businesses, vertical or horizontal integration through acquisition, financial engineering, and company turnaround. Buyout managers need to give extensive advice on strategic and business planning, and they tend to focus on consistent rather than outsized returns. Because they target established enterprises, buyout firms experience fewer outright failures but have more limited upside potential.

7.2.2 Deal Structuring of Venture Capital and Buyout

VC transactions do not typically involve debt, but venture capitalists gain control of a company over time through a series of equity investments. Returns stem from building companies and from managing growth. Valuation is complicated by the lack of appropriate comparisons, which explains why venture capitalists carry out more extensive sector/product due diligence and more limited financial due diligence compared to buyout managers. Venture capitalists typically provide not only financing for building businesses but also industry know-how, relevant contacts, and management expertise. The investments can be relatively small and are overwhelmingly equity or quasi-equity financed, with little or no leverage.[3] Successful exit strategies require VC managers to secure follow-on financing.

Buyout transactions, on the other hand, typically use both equity and debt financing to acquire companies. Assets of the acquired company are used as collateral for the debt, and the company's cash flow is used to pay off the debt. Buyout managers conduct intensive financial due diligence and occasionally rely on sophisticated financial engineering. Financial engineering refers to the process of creating an optimal capital structure for a company. In private equity, the capital structure is often made up of different types of financial instruments, such as multiple layers of debt, mezzanine, and equity, each carrying a different risk-reward profile. The ability to analyze a company's balance sheet and extract operational efficiencies, as opposed to the implementation of capital structure changes, is the primary driver of a successful transaction. Generally, there are few limitations to investment size, given the high number of both privately held and publicly traded stable-growth and mature companies that can be targeted.

7.2.3 Role of the PE Manager in Venture Capital and Buyout

Depending on the strategy, the role of the PE manager can differ dramatically. Venture capitalists look to launch new or emerging companies, whereas buyout managers focus on leveraging an established company's assets. Venture capitalists back entrepreneurs, whereas buyout managers deal with experienced managers. Venture capitalists often play an active role in the companies in which they invest, by either sitting on the board of directors or becoming involved in the day-to-day management of the company.

In buyout transactions, a greater proportion of time and manpower is spent analyzing specific investments and adjusting the business model. Buyout managers look to leverage their expertise in order to turn around underperforming businesses, improve profitable businesses, or optimize the companies' balance sheets and financing. They typically engage in hiring new management teams or retooling strategies. In an operating company, it is easier to give guidance to a seasoned management team, whereas in early-stage investments, one often needs to build and coach the management team from the ground up.

7.3 PE FUNDS AS INTERMEDIARIES

PE funds serve as financial intermediaries by facilitating investment in private equity. There are several motivations to using PE funds as intermediaries.

7.3.1 PE Fund Intermediation and Risk

Basically, private equity funds step into the funding process when traditional lenders are not willing or able to provide funding. For example, banks may be unwilling to lend to an entrepreneur or be involved in leveraged buyouts because of the significant risk that is involved. In the case of lending to an entrepreneur, the product or the intellectual property is not well understood, and of course the firm has no track record that the bank could use to evaluate its riskiness. Also, some of these lenders (e.g., banks) are not willing or allowed to take equity positions in these firms. This means they cannot fully participate in the significant upsides that PE investments could provide. These significant upside returns can justify the risks PE investors are willing to take by investing in new and untested companies or by taking significant credit and leverage risk and investing in poorly performing firms. PE investors take advantage of the inefficiencies in financial markets to satisfy the needs of these borrowers.

7.3.2 PE Fund Intermediation and Efficient Incentives

Another reason for the existence of PE firms is the presence of certain economic inefficiencies in the traditional corporate structure. In some cases, management may not be given proper incentives to maximize the value to the shareholders of a corporation. Private equity seeks to address this problem by tightly aligning the interests of managers and shareholders to achieve increased efficiency and higher return to shareholders.

PE funds principally serve the following functions:

- Pooling of investors' capital for investing in private companies
- Screening, evaluating, and selecting potential companies with expected high-return opportunities

- Financing companies to develop new products and technologies, foster their growth and development, make acquisitions, or allow for a buyout or a buy-in by experienced managers
- Controlling, coaching, and monitoring portfolio companies
- Sourcing exit opportunities for portfolio companies

This is a classic principal-agent (limited partner-general partner [LP-GP]) relationship, which, because information in PE markets is incomplete and highly asymmetric, requires some specific agreements to cover the resulting problems of moral hazard and conflict of interest. While the specific terms and conditions plus investor rights and obligations are defined in nonstandard partnership agreements, the limited partnership structure (or comparable structures used in various jurisdictions) has evolved over the last decades to include the following standards:

- The fund usually has a contractually limited life of seven to 10 years, often with a provision for an extension of two to three years. The fund manager's objective is to realize, or exit, all investments before or at the liquidation of the fund.
- Investors, mainly institutions like pension funds, endowments, PE funds of funds, public institutions, banks, insurance companies, and high-net-worth individuals or family offices, are the LPs and commit a certain amount of money to the fund.
- Commitments (capital pledges by investors in PE funds) are drawn down as needed, or just in time, to make investments or to pay costs, expenses, or management fees. Because PE funds do not typically retain large pools of uninvested capital, their GPs make capital calls (or **drawdowns**) once they have identified a company in which to invest. Therefore, the main part of the drawdown gets invested immediately.
- Generally, the year in which the fund first draws down capital from investors for the purpose of investing in a company is called the vintage year; some data providers use the year that the fund commences operations as the vintage year.
- A significant portion, though not typically all, of committed capital is drawn down during the investment period, typically the first three to five years, during which new opportunities are identified. After that, during the divestment period, only the existing portfolio companies with the highest potential are further supported, with some follow-on funding provided to extract the maximum value through exits. The manager's efforts during this time are concentrated on realizing or selling investments.
- When **realizations** (sales of portfolio companies) are made, or when interest payments, dividends, or recapitalizations are received, they are distributed to investors as soon as is feasible. Funds may have a reinvestment provision, wherein the proceeds of realizations within the investment period or a similar time frame may be reinvested in new opportunities and not distributed to investors. Under this scenario, the fund is self-liquidating as the underlying investments are realized. However, these returns come mostly in the second half of the fund's lifetime.
- These distributions to investors can also take the form of securities of a portfolio company, known as **in-kind distributions**, provided that these securities are publicly tradable or distributed when the fund gets liquidated. Legal documentation

may also allow for some reinvestment of realizations, normally subject to a cap amount.

- Management fees depend on the size of the fund, generally ranging from 2.5% of committed capital for small funds to 1.5% for larger funds. The fees are often based on the amount of committed capital during the investment period and on the value of the portfolio thereafter. There are considerable differences from one fund to the next regarding directorship fees and transaction costs. These can have an impact on returns and often account for material differences between gross and net returns.
- The main upside incentive for GPs comes in the form of carried interest, typically 20% of the profits realized by the fund. Carried interest is usually subject to a hurdle rate, or preferred return, so that it begins to accrue only once investors have received their capital back and a minimum previously agreed-on rate of return. Once the preferred return has been attained, GPs typically receive 100% of returns to the point at which they would have received the carried interest on the entire amount. This is called a catch-up and is synonymous with the soft hurdle concept used by hedge funds.

There is a PE fundraising cycle that begins anew each time GPs need to raise capital for another fund. Typically, limited partnership agreements do not allow follow-on funds by the same manager before the end of the initial fund's investment period or until a large part of the initial fund has been invested.

7.3.3 Forms of PE Fund Intermediation

There are different routes for investing in private equity (see Exhibit 7.2). Few institutions have the experience, the incentive structures, and the access that would allow them to invest directly in nonpublic companies, so most investors seek intermediation

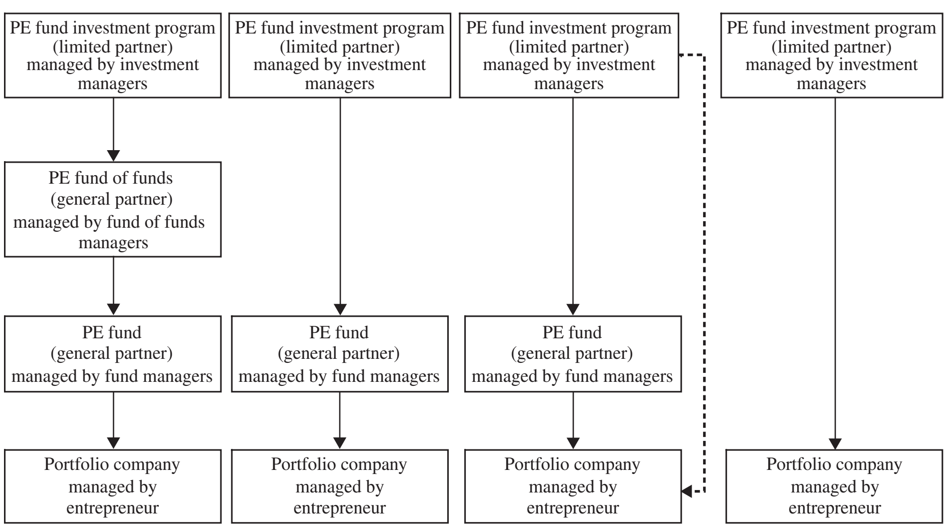


EXHIBIT 7.2 PE Funds Investment Program

through funds. For institutions, the most relevant approaches to investing in private equity are through fund-of-funds specialists as intermediaries or through similarly structured dedicated in-house PE investment programs that invest directly in funds. Other routes are via publicly quoted PE vehicles or through a dedicated account managed by a PE specialist.

Reading Exhibit 7.2 from left to right, the various programs are defined as follows:

- In a fund of funds structure, the PE fund investment program buys units of a PE fund of funds' general partner, who in turn purchases units of a PE fund, which further invests in a portfolio company.
- A PE fund is more direct in that the investment is into a PE fund and then into a portfolio company.
- A PE fund with co-investment adds a co-investment leg, wherein the PE fund has an additional investment in a certain portfolio company, typically at preferential management and performance fee terms. **Co-investment** refers to the practice of investors being invited by GPs to make direct investments in portfolio companies. Co-investment is discussed later in this chapter.
- **Direct investment in private equity** eschews PE funds altogether, as the PE investment program makes investments straight into a portfolio company (without intermediation), similar to a co-investment but without the input of a PE fund manager.

7.3.4 The Markets for PE Funds

The organized PE market is dominated by funds, generally structured as limited partnerships, which serve as principal financial intermediaries. Fund management companies, also referred to as PE firms, set up these funds. PE funds are unregistered investment vehicles in which investors, or **limited partners (LPs)**, pool money to invest in privately held companies. Investment professionals, including venture capitalists and buyout managers, manage these funds and are known as **general partners (GPs)** or fund managers. Tax, legal, and regulatory requirements drive the structuring of these investment vehicles with the goal of increasing transparency (investors are treated as investing directly in the underlying portfolio companies), reducing taxation, and limiting liability (investors' liabilities are limited to the capital committed to the fund). From a strictly legal standpoint, limited partnership shares are illiquid; in practice, however, **secondary transactions in private equity** occasionally take place, in which investors sell their shares to other investors before the termination of the fund.

7.4 PE FUNDS OF FUNDS AS INTERMEDIARIES

Many institutions outsource their PE fund investment program either through a dedicated account or by pooling assets with other investors. PE funds of funds are probably the most common type of institutional investment program. The authorizing entity for a PE fund investment program is the principal who provides the resources, while the manager of the program is the principal's agent and conducts the investments in PE funds as an LP.

PE funds of funds, which are mainly organized by specialist asset managers, are vehicles that pool capital from a group of investors to invest in a diversified portfolio of funds. Some funds of funds specialize in certain PE sectors or geographies, whereas others follow a more generalist approach. Funds of funds manage the following, often complementary, activities:

- Primary investments in newly formed limited partnerships. Because of the **blind-pool** nature of such investments, in which investors don't know the underlying portfolio companies before committing capital, the assessment of the fund management team's skills is key (see Chapter 9).
- Co-investments alongside primary investments. This activity requires direct investment experience and skills.
- Secondary investments in existing funds or portfolios of direct investments. This is generally a niche activity for most funds of funds; however, secondary specialists, such as Coller Capital, Greenpark Capital, and Lexington Partners, are major players in this market. This activity requires both co-investment skills for the assessment of the companies already in the portfolio and primary investment skills for the blind-pool part of the transaction.

While investment in a particular PE fund can have a blind-pool nature, a fund of funds can have established relationships with fund managers via existing investments. Therefore, its future portfolio is somewhat predictable and is not necessarily a blind-pool investment. A newly created portfolio is likely to be largely composed of follow-on funds raised by these known managers. In fact, funds of funds are marketed on either a partially blind or a fully informed basis. For a partially blind pool, some of the intended partnership groups are identified, while for a fully informed pool, virtually all of the intended partnerships have been identified.

7.4.1 PE Funds-of-Funds Costs

Funds of funds are often seen as less efficient than single GP funds (or direct fund investments) because of the additional layer of management fees. This double layer of fees is perceived to be one of the main disadvantages of this structure. Funds of funds would have to outperform direct fund investment to compensate for this additional layer of fees.[4] However, given the resources required to manage a portfolio of PE funds internally, investing through a fund-of-funds structure might well prove more cost-efficient than direct fund investments in the end.

An additional cost of outsourcing to a fund of funds is the carried interest. Whether an in-house program can work without investment performance–related incentives is debatable. According to Otterlei and Barrington (2003), the annual costs of an in-house team can be significant compared to that of a typical fund of funds. Even with a 5% carried interest charged by the fund-of-funds manager, these authors find that the fees have an insignificant impact on the net returns of the investor. However, information is an asset in the often opaque environment of private equity. Taking the fund-of-funds route versus that of direct investor can lead to a loss of information and control, essentially a cost in itself.

Because PE programs follow a learning curve, inexperienced institutions may initially have little option other than to go through a fund-of-funds vehicle. Ultimately, they can become LPs in funds and, with increasing sophistication, build their own portfolios of companies, either through co-investing or by independently sourcing deals. In conclusion, funds of funds are often used as a first step into private equity and may well be worth the additional layer of fees in exchange for avoiding expensive learning-curve mistakes and providing access to a broader selection of funds.

7.4.2 Added Value of PE Funds of Funds

Investing in funds of funds can allow investors access to the PE market in a quick and diversified manner. Before making such an allocation, there are several factors that investors need to consider.

7.4.2.1 Diversification and Intermediation Funds of funds can add value in several respects and are seen as safe havens for PE investors. Especially in the case of new technologies, new teams, or emerging markets, a fund of funds allows for reasonable downside protection through diversification. Not surprisingly, various studies have shown that because of their diversification, funds of funds perform similarly to individual funds but with less pronounced extremes (see Weidig and Mathonet 2004; Mathonet and Meyer 2007). In the absence of funds of funds, smaller institutions may have difficulty achieving meaningful levels of diversification. Even for larger institutions, investments in PE funds and especially VC funds may be too cost-intensive when the size of such investments is small compared to administrative expenses. A fund of funds can mediate these potential size issues by either **scaling up** through pooling of commitments of smaller investors and providing each of them with sufficient diversification, or **scaling down** through sharing administrative expenses and making such investments less cost-intensive by allowing larger commitment to the fund of funds.

7.4.2.2 Resources and Information Funds of funds can provide the necessary resources and address the information gap for inexperienced PE investors through their expertise in due diligence, monitoring, and restructuring. Investing in PE funds requires a wide-reaching network of contacts in order to gain access to high-quality funds, trained investment judgment, and the ability to assemble balanced portfolios. Liquidity management can also be quite challenging, as it demands a full-time team with insight and an industry network, adequate resources, and access to research databases and models, as well as skills and experience in due diligence, negotiation, and contract structuring. Depending on the overall market situation, access to quality funds can be highly competitive, and being a newcomer to the market can pose a significant barrier. Funds of funds are continuously involved in the PE space, speak the language, and understand the trade-offs in the industry.

7.4.2.3 Selection Skills and Expertise Investors expect funds-of-funds managers to be able to invest in top-performing funds, either by having access to successful invitation-only funds or by identifying the future stars among the young and lesser-known funds. Funds-of-funds managers may also play the role of educator in explaining to comparatively unsophisticated investors that a particular fund, despite suffering horrible losses in the early years, is still viable and merely reflecting the early stages of its J-curve (see Chapter 8). While funds of funds are more willing to give fund managers sufficient latitude to focus on their portfolio companies, they are often more skilled and experienced in restructuring failing funds, if that is ultimately required. In turn, fund managers often welcome funds-of-funds investors as a more stable and experienced source of pooled capital.

7.4.2.4 Incentives and Oversight For institutional investors, direct investment is problematic because such institutions often cannot offer their employees adequate performance-related pay. For typical conservative and seniority-based institutions like banks, pension funds, and insurance companies, a theoretically unlimited carried interest does not always fit well into the compensation scheme. While institutional investors do not lack staff with the intellectual caliber to evaluate investment proposals and to structure transactions, generating profitable exits in PE programs requires very hard work over protracted periods of time. Moreover, the lack of incentive to take risk and to find value (or the conflict of interest therein) may affect investment decisions. Furthermore, there is a significant learning curve, and without performance-related pay, employees may jump ship as soon as they are competent in the area and understand their opportunities better. Finally, for larger institutions, intermediation through funds of funds allows them to focus on their core businesses. This advantage tends to outweigh most cost considerations.

7.5 THE RELATIONSHIP LIFE CYCLE BETWEEN LPs AND GPs

There is a symbiotic relationship between LPs and GPs. An LP's investment strategy is built around a number of relationships with GPs, who focus on specific segments (such as stages or sectors) of the market. This specialized focus can often limit the scalability of a particular fund, especially in the case of venture capital, in which LPs may find it difficult to identify and access additional fund managers of comparable quality.

GPs, for their part, want financially strong, dependable, knowledgeable, and long-term LPs. LPs should have industry expertise and familiarity with the nuts and bolts (particularly valuations and benchmarking) of the PE business. Adverse selection exists in the PE market. Poor-quality GPs, be they lacking experience or falling into decline, will court inexperienced LPs. Because of poor results, both will eventually exit the market.

To maintain continuous investment in new portfolio companies, GPs need to raise new funds as soon as the capital from their latest active fund is fully invested (or reserved for follow-on investments), that is, about every three to five years. Thus, relationships between LPs and GPs follow a life cycle and are forged through various rounds of investment, eventually resulting in a virtuous circle of growing experience and fund size.

Anecdotal evidence suggests that experienced market players profit from these relationships over protracted periods of time. Initially, criteria are very stringent, and fund managers usually cannot get rich through their first funds. However, a favorable track record is an asset in itself. For more reputable funds, fundraising is less costly.

To minimize their expenses, fund managers generally turn first to those who invested in their previous partnership, provided that the fund's performance was satisfactory.

While it is easy to see how fund managers benefit from a loyal and reliable investor community, these long-term relationships can also be advantageous for LPs for several reasons:

- In the opaque PE market, the search for and due diligence of funds is a costly exercise, and LPs often prefer familiar fund managers to unproven investment proposals.
- Such long-term relationships may provide access to a quality deal flow of co-investment opportunities in portfolio companies within an established framework.
- It is especially desirable for an investor to hold on to good fund managers, as the best teams will have an established investor base (i.e., a set of established and loyal clients), which may eliminate the need to seek out new funding sources to the detriment of adding value to the portfolio companies when making new investments or exits.
- There is likely to be better planning, as LPs make clear their intentions to participate in follow-on funds. As LPs form a network, even if they do not have the means to continue, they often refer other investors to a good team. Predictable closings put money to work more efficiently.

The **life cycle of the GP–LP relationship** (see Exhibit 7.3) focuses on the long-term pattern of GPs as they create multiple funds through time. The GP–LP life cycle can be divided into three phases: (1) **entry and establish** (the phase involving the initial funds); (2) **build and harvest** (or grow and compete, the phase in which the funds thrive and grow); and (3) **decline** (lost competition), **exit** (gave up or made it), or transition to new managers (**spinouts**), the three potential outcomes leading

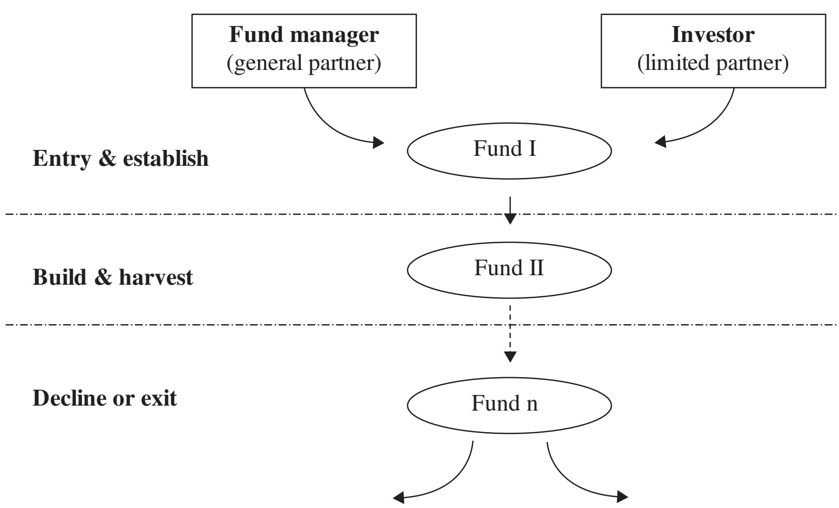


EXHIBIT 7.3 Fund Manager–Investor Relationship Life Cycle

EXHIBIT 7.4 GP–LP Relationship Life Cycle Model

Fund Characteristic	Entry and Establish	Build and Harvest	Decline or Exit
Investment strategy	Differentiation	Star brand	Unexciting
Fundraising	Difficult fundraising	Loyal LP base	LPs leave and are replaced by other types of investors (secondary plays, new entrants in market)
Performance	Unknown: either top or out	Likely top performer	Not top but consistent performer
Size	Fund is too small	Fund size is right	Fund size too large/too many funds
Economies of scale	Fund is too small to get rich	Best alignment of interests	Senior managers made it
Management team	Management team forming	Management team performing	Succession issues, spinouts

toward the termination of the original relationship. The main differences between these phases are summarized in Exhibit 7.4.

During the entry and establish phase, substantial entry barriers into the PE market exist for both GPs and LPs. Lacking a verifiable track record, new teams find it difficult to raise their first fund. Furthermore, analysis of historical benchmark data supports the hypothesis that new teams suffer from higher mortality than do established or institutional-quality fund managers. First-time funds note the importance of differentiation or innovation as applied to fundraising and thus often pursue specialized investment strategies.

New LPs also face entry barriers, suffering the initial informational disadvantages that make it extremely difficult to identify or gain access to the best managers, particularly when their funds are oversubscribed. For LPs, it takes the disciplined execution of a long-term investment strategy to build up a portfolio of funds that gives attractive and sustainable returns.

Since investors are mainly interested in the cash returned, the fund manager–investor relationship tends to be relatively stable throughout the build and harvest phase. Lerner and Schoar (2004) present evidence on the high degree of continuity in the investors of successive funds, and the ability of sophisticated investors to anticipate funds that will have poor subsequent performance.

It is an oversimplification to assume that investors invest only in top performers and that below-average funds are unable to continue. As in most relationships, there is a certain degree of tolerance for mistakes and failures, at least over a period of time. It is clear that there are limits to disappointing results, but all things being equal, investors will tend to go with fund managers they already know or who have been referred to them through their network, even if the fund's performance has been subpar at times.

Eventually, the relationship ends in the decline, exit, or spinout phase. Not surprisingly, the terms *marriage* and *divorce* are often used in the context of relationships between fund managers and their investors. A gradual decline may occur either as a result of past successes, which potentially decrease the financial motivation of senior fund managers, or due to an improperly planned succession, which leads to the departure of middle management. In addition, the LPs may eventually end the relationship if they lose confidence or trust in the team (for example, if the team becomes arrogant or fails to deliver). Some LPs do not invest in follow-on funds and may be replaced by less deep-pocketed or experienced investors, or by secondary investors who choose to invest as a one-off financial play.

7.6 LIMITED PARTNERSHIP KEY FEATURES

For the PE fund, the **limited partnership agreement (LPA)** defines its legal framework and its terms and conditions. The LPA has two main categories of clauses: (1) investor protection clauses, and (2) economic terms clauses. Investor protection clauses cover investment strategy, including possible investment restrictions, key-person provisions, termination and divorce, the investment committee, the LP advisory committee, exclusivity, and conflicts. Economic terms clauses include management fees and expenses, the GP's contribution, and the distribution waterfall. The **distribution waterfall** defines how returns are split between the LP and GP and how fees are calculated. LPAs are continuously evolving, given the increasing sophistication of fund managers and investors, new regulations, and changing economic environments. In essence, the LPA lays out conditions aimed at both aligning the interests of fund managers with their investors and discouraging the GP from cheating (moral hazard), lying (adverse selection), or engaging in opportunism (holdup problem) in whatever form.

Moral hazard and adverse selection take place when there is asymmetric information between two parties (e.g., LPs and GPs). **Adverse selection** takes place before a transaction is completed, when the decisions made by one party cause less desirable parties to be attracted to the transaction. For example, if an LP decides to seek GPs that charge very low fees and offer funds with very favorable terms, the LP is likely to attract unskilled GPs that claim to be skilled.

Moral hazard, in contrast, takes place after a transaction is completed and can be defined as the changes in behavior of one or more parties as a result of incentives that come into play once a contract is in effect. For example, without proper monitoring, a GP may take excessive risk in order to increase the potential performance fee, or an unskilled manager may decide not to make any investment and just collect the management fee.

In economics, the holdup problem is a situation in which two parties (in the case of private equity, a GP and an LP) refrain from cooperating due to concerns that they might give the other party increased bargaining power and thereby reduce their own profits. Incentives are designed so that the fund manager's focus is on maximizing terminal wealth and performance, and ensuring that contractual loopholes are not exploited (e.g., by producing overly optimistic interim results).

7.6.1 Fit with Existing Market Environment

In its current form, the limited partnership is viewed as a good fit for the existing market environment, yet the environment is continuously evolving. However, given the success of the current limited partnership structure, it is more likely we will see adjustments to the LPA terms and conditions rather than broad structural changes. To some degree, an increasing standardization can be observed, as illustrated by the increasing adoption of the Private Equity Principles put forward by the Institutional Limited Partners Association (ILPA).[5] But evolution still exists, notably due to the PE market's high degree of fragmentation, competition, and variations in local regulations, which can precipitate different fund terms and conditions in the LPA.

It is through the proper alignment of the economic interests of investors and managers, not just through the LPA covenants, the advisory boards, or the committees composed of LPs, that one can eliminate many of the problems associated with the principal-agent relationship, especially in those scenarios that cannot be foreseen. To be successful, the structure must address management fees, performance-related incentives, hurdle rates, and, most importantly, the fact that GPs make a substantial personal investment in each fund. Additional clauses may be required to cover reinvestments and clawbacks, as well as noneconomic terms such as key-person provisions, joint and several liability, and disclosure obligations. Together, these clauses provide LPs with moderate but sufficient control over the management of the fund. These key LPA features are described in the following sections.

7.6.2 Corporate Governance in PE Funds

The law and the LPA define and restrict the degree of control LPs have over the activities of GPs. Such controls relate, for example, to waiving or accepting investment restrictions, extending the investment period or fund duration, handling key-person-related issues, or participating in an **LP advisory committee (LPAC)**, whose responsibilities are defined in the LPA and normally relate to dealing with conflicts of interest, reviewing valuation methodologies, and any other consents predefined in the LPA. LPs can make decisions with either a simple majority (e.g., the decision to extend the investment period or the fund's duration) or a qualified majority (e.g., the decision to remove the GP without cause). A **qualified majority** is generally more than 75% of LPs as opposed to the 50% required for a simple majority.

Occasionally, LPs may be offered positions on the investment committee. However, it is not clear whether LPs should actually take on this role. In limited partnership structures, an overactive LP could become reclassified as a GP, thereby losing his or her limited liability. Generally, international industry professionals recognize that fund managers should make investment and divestment decisions without the direct involvement of investors, so as not to dilute the responsibility of the manager, create potential conflicts of interest with nonparticipating investors, or expose LPs to the risk of losing their limited liability. Also, investors do not normally have the legal rights or the required skills and experience to make such decisions.

Another important element of corporate governance is reporting to LPs. Various PE associations or industry boards, including the European Private Equity and Venture Capital Association (EVCA), the Institutional Limited Partners Association (ILPA), the International Private Equity and Venture Capital Valuation Board (IPEV Valuation Board),[6] and the Private Equity Industry Guidelines Group (PEIGG),[7] have released guidelines for valuation and reporting. The obligation to disclose in compliance with these guidelines is increasingly being made part of contractual agreements. While some GPs reduce the level of detail provided to the bare minimum and share it with all LPs, others share different levels of detail depending on the specific type of investor.[8]

7.6.3 Investment Objectives, Fund Size, and Fund Term

In LPAs, the description of investment objectives should be specific but not too narrow. Lerner (2000) argues that PE funds are blind pools for a reason. Investors should not attempt to put overly restrictive limits on a fund manager's flexibility, which could block the fund manager's ability to profit from unanticipated opportunities. Further, with uncertain investments, severe information asymmetry, and difficulties in monitoring and enforcing restrictions, fund managers may simply find ways around narrow restrictions.

The fund's size, in terms of capital committed by LPs, needs to be in line with these investment objectives. However, various factors, such as the management resources required or the number of potential opportunities, implicitly set a minimum or maximum size of the fund.

Fund lives or terms are typically seven to 10 years, with possible extensions of up to three years. This represents a trade-off among better investment returns, sufficient time to invest and divest, and the degree of illiquidity still acceptable for investors. Normally, the extension of a fund's life is approved annually by a simple majority of LPs or members of the LPAC, one year at a time versus two or more at once, during which time management fees are either reduced or eliminated altogether to stimulate quick exits.

Normally, proceeds are distributed to investors as soon as is feasible after the realization of or distribution of a fund's assets, but in some cases, LPs grant fund managers the discretion to reinvest some of the proceeds that are realized during the investment period.

7.6.4 Management Fees and Expenses

In private equity, compensation is overwhelmingly performance driven. Management fees provide a base compensation so that the fund manager can support the ongoing activities of funds, and there is a consensus that GPs should not be able to make significant profits on management fees alone. These fees need to be based on reasonable operating expenses and salaries, and be set at a level modest enough to ensure that the fund manager is motivated primarily by the carried interest, but sufficient to avoid the manager's departure to greener fields. During the investment period, fees are on a commitment basis, rather than being levied only against capital that has been invested. If fees were based only on invested capital, management might have an incentive during the investment period to pursue volume instead of quality. Also, the fees on committed but uninvested funds can be viewed as paying for the staffing required for due diligence during the investment period. At the end of the investment period or with the raising of subsequent funds, fees ramp down and are levied only against capital still under management. Fees are adjusted according to the proportion of the portfolio that has been divested.

When management fees are low, the fund manager may be looking for other compensation, such as that earned by sitting on the boards of directors of portfolio companies, providing advisory or management services, or advising on or structuring transactions. To counter these incentives for distraction from investment management or the double payment of fund management services, a **fee-offset** arrangement may be implemented, in which fees are fully or partially netted against management fees. Fee-offset clauses have become standard in LPAs.

7.6.5 Carried Interest

Carried interest, or carry, is the share in the fund's profits received by the fund manager. Carried interest should be predominantly directed to the professional staff responsible for the success of the fund, and vested in parallel with the expected value-creation time line. It remains largely uncontested over time that the fund manager's main incentive is performance based through the realization of carried interest. Carried interest is calculated on either a fund-as-a-whole or a deal-by-deal basis. In both cases, the carried interest is normally payable if the internal rate of return (IRR) of a fund exceeds the hurdle rate or preferred return (see section 7.6.6). Using fund-as-a-whole accounting, previous losses and overall portfolio performance are also taken into account. It is generally accepted that a carried interest of 20% aligns manager and investor interests appropriately. Typically, LPs favor the fund-as-a-whole over the deal-by-deal methodology for carried-interest calculations, as it further aligns manager and investor interests.

Since the beginning of the Great Recession, LPs have been overwhelmingly rejecting deal-by-deal carry. But whether this approach is now really extinct or the pendulum will swing back toward more GP-friendly carried interest mechanisms remains to be seen.

In order to compare these two methods for calculating carried interest, we may consider the simplified scenario shown in Exhibit 7.5, in which LPs contribute in the first year €100 million to fund investments A and B, at €50 million each, with an 80/20 carry split.

Investment A is sold in the second year for €90 million. The profits of €40 million for this investment are distributed to LPs and the GP in line with the agreed-on 80/20 split. In the third year, investment B fails and is written off entirely.

The fund as a whole has a loss of €10 million and gives no right to carried interest for the GP, while under the deal-by-deal approach, the GP would receive €8 million of carried interest and the LPs would have a loss of €18 million.

7.6.6 Preferred Return

The preferred return, often also called the hurdle rate, has become a standard limited partnership term:

- The typical explanation for its introduction is to ensure that GPs are compensated only for outperformance (Maxwell 2003). Therefore, LPs have first priority to receive all distributions up to their commitment (or capital invested), plus the preferred return, before the GP gets access to carried interest. In theory, then, the preferred return is an annual compounded interest rate on the invested capital that is set well above the currently prevailing risk-free rate of return, but at or below the historical performance of public equity. If the hurdle rate is set too low, it becomes meaningless as an incentive and just creates administrative problems.

EXHIBIT 7.5 Illustration of Deal-by-Deal Carried Interest

Year 1, Deal-by-Deal

	LPs Investment A	Investment B	GP	Total
	–€50 million	–€50 million		–€100 million
Acquisition of investments A and B				
Closing balance	–€50 million	–€50 million		–€100 million

Year 2, Deal-by-Deal

	LPs Investment A	Investment B	GP	Total
Opening balance	–€50 million	–€50 million	€0 million	–€100 million
Sale of investment A for €90 million				
Return of capital	€50 million			€50 million
80/20 split of residual amount	€32 million		€8 million	€40 million
Closing balance	€32 million	–€50 million	€8 million	–€10 million

Year 3, Deal-by-Deal

	LPs Investment A	Investment B	GP	Total
Opening balance	€32 million	–€50 million	€8 million	–€10 million
Total write-off of investment B				
Return of capital		€0 million		€0 million
80/20 split of residual amount				
Closing balance	€32 million	–€50 million	€8 million	–€10 million
Subtotal	–€18 million		€8 million	–€10 million

- While managers of buyout funds generally offer their investors an 8% preferred return before they take a share of the profits for themselves, this is usually not the case for venture capitalists, at least in the United States. In this context, Fleischer (2005) puts forward an alternative explanation for the preferred return: it is more important as an incentive to properly screen the deal flow. Without it, buyout fund managers could pursue a low-risk, low-return strategy, for example, by being inactive or by choosing companies that have little potential to generate large returns. A preferred return forces managers to make riskier investments to generate a return in line with the investor's targets. In the case of VC funds, however, investments are always risky, and the high-risk, high-reward strategy makes it meaningless to bother about preferred returns.

Though high hurdle rates aim to provide an incentive for fund managers to outperform, they can also have the opposite effect. Managers of funds with overly high hurdle rates (or of struggling funds) can be demotivated if it becomes unlikely that they will receive carried interest. High hurdle rates may also lead to excessive risk taking.

There are also some conceptual issues with preferred returns. GPs are faced with the dilemma of whether to realize an investment over a short period of time to optimize the IRR or to hold on to it and try to optimize the multiple. For example, is it better to generate a 50% IRR for a period of three months, which yields a 1.11× multiple on capital invested, or only a 10% IRR for a period of three years, leading to a 1.33× multiple? The standard preferred return, being based on the IRR, gives incentive to the former. An alternative is to base the preferred return on the multiple.

Incentive fees are used to reward managers when the fund does well, but the absence of these fees alone does not provide sufficient punishment for when the manager underperforms. With this kind of structure, the more risk a manager takes, the greater the upside potential, with little immediate downside impact from losses. This is why the GP's capital contribution is so important.

7.6.7 GP's Contribution

Excessive risk taking can be reduced or eliminated if managers have a significant portion of their personal wealth in the fund. In this case the manager, being directly exposed to fund losses, no longer has an incentive to take excessive risks, to work on non-fund-related activities, or to abandon ship once the prospects for generating carry and launching a follow-on fund become highly unlikely.

Typically, investors in PE funds see 1% of the fund's committed capital as a standard and acceptable contribution for GPs to make. This capital contribution by GPs, also known as **hurt money**, should be contributed in cash rather than through the waiver of management fees (or as surplus from the management company's budget). However, in the case of wealthy managers, 1% may at times be too low.[9] To better understand this relationship, it makes sense to look at the GP's and LPs' relative exposure. Typically, the GP's contribution to the fund is a significant share of his or her personal wealth, whereas the LPs' investment, although in absolute terms far higher, represents an immaterial share of the institution's overall assets. It is a challenge to determine the appropriate contribution level for a GP: one that provides a reasonable incentive but is not excessively onerous. An analysis of profits earned from past investments, salaries, budget surpluses, and so forth may provide useful information for determining a contribution level that is appropriate for a particular GP.

7.6.8 Key-Person Provision

Private equity is to a large degree a people business. Consequently, the key-person provision is a common and important clause. Depending on the size, experience, and depth of the team, the inability of one or several key persons to carry out their duties could have a substantial adverse effect on the partnership.

If one or several of the named key persons depart the team, stop committing sufficient time to the management of the fund, or sell their interests in the management company, the key-person provision allows LPs to suspend investment/divestment activities until a replacement or replacements are found. The LPs can even terminate the fund if they so choose. Key-person clauses may also be put in place in anticipation of the retirement of senior fund managers.

7.6.9 Termination and Divorce

LPAs may foresee a for-cause removal of the GP and include a **bad-leaver clause**, which, if exercised (normally following a simple majority vote of the LPs), causes investments to be suspended until a new fund manager is elected or, in the extreme, the fund is liquidated. In practical terms, conditions leading to a for-cause removal are difficult to both define and determine. In private equity, very little can be legally enforced; issues are highly subjective, and taking matters to court carries high legal risk for an investor, as it is very difficult and lengthy to prove wrongdoing.

The **good-leaver clause** enables investors to cease additional funding of the partnership with a vote requiring a qualified majority, generally more than 75% of LPs. This "without-cause" clause provides a clear framework for shutting down a partnership that is not working, or when confidence is lost. The good-leaver clause sometimes provides for compensation amounting to six months to one year of management fees; the bad-leaver clause provides no such compensation and foresees no entitlement to carried interest. This contrasts with the good-leaver clause but includes a vesting schedule, so that part of the carry remains available to incentivize the new team being hired.

In deciding whether to exercise one of these clauses, reputation considerations play a key role, as the market consists of a small number of players who repeatedly interact with one another. GPs who are removed with or without cause may subsequently be unable to raise funds or participate in investment syndicates with other partnerships. To avoid such disastrous outcomes, GPs tend to agree on a fund restructuring to prevent a forced removal.

7.6.10 Distribution Waterfall

The EVCA glossary defines exit as the "liquidation of holdings by a PE fund. Among the various methods of exiting an investment are: trade sale; sale by public offering (including IPO); write-offs; repayment of preference shares/loans; sale to another venture capitalist; sale to a financial institution." Proceeds from these exits are split in terms of amount as well as timing between GPs and LPs according to the distribution waterfall.

Normally, returned to LPs are, first, all the capital invested plus fees and expenses (i.e., all drawn capital), or sometimes the total commitment (both drawn and undrawn capital). They then receive the preferred return on the investment (i.e., the hurdle zone), normally compounded per annum (see Exhibit 7.6). This is followed by the catch-up period, during which the GP receives all or the major share of distributions. The catch-up period ends when the agreed-on carried-interest split, or stated share of the fund's profits to be received by the GP and the LPs, is reached. Thereafter, distributions are shared between the fund manager and investors in the proportion agreed to in the legal documentation, normally 20% for the GP and 80% for LPs. This mechanism is also not standardized (e.g., whether an interest rate such as a hurdle rate is compounded on a quarterly or on an annual basis). Sometimes there is a full catch-up, whereas other agreements may foresee only a partial catch-up or no

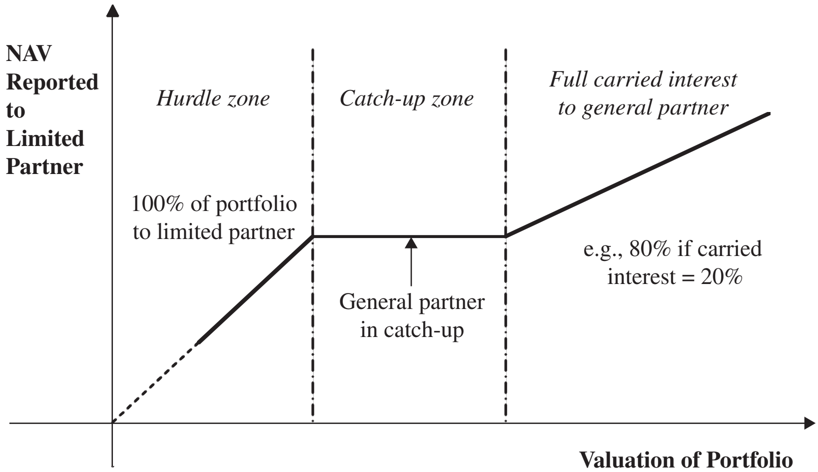


EXHIBIT 7.6 Impact of Catch-Up on NAV Attributed to LPs

catch-up at all. With a 100% catch-up, the LPs are effectively excluded from all new distributions until the GP has fully recovered the agreed-on GP share of the profit.

To put it in more mathematical terms, the function that maps the gross asset value onto the net asset value (NAV) is not information preserving. In situations in which there is a 100% catch-up, several different gross asset values, representing the underlying portfolio's value, will be mapped to the same NAV due to the fund's LPs.

As a simplified example of the basic waterfall, we assume a €100 million contribution by the LPs in the first year to fund an investment, an 8% hurdle rate, a 100% catch-up, an 80/20 carry split, and the sale of the investment by the fund in the second year for €200 million. In this case, the sales proceeds are distributed as shown in Exhibit 7.7.

Note that simply splitting the profit of €100 million (the difference of €200 million in sales proceeds minus the €100 million purchase price) on an 80/20 basis would have given the same result. The reason for this is that the fund's return is above the catch-up zone, and therefore the hurdle is extinguished and does not make any difference.

EXHIBIT 7.7 Example of Waterfall Using a Hurdle Rate

	LPs	GP	Total
Original contribution	–€100 million		–€100 million
Sale of investment for €200 million			
Return of capital	€100 million		€100 million
Preferred return for limited partners	€8 million		€8 million
Catch-up for GP		€2 million	€2 million
80/20 split of residual amount	€72 million	€18 million	€90 million
Closing balance	€80 million	€20 million	€100 million

The simplest and, from the viewpoint of an LP, most desirable solution is that GPs do not take carried interest until all invested, or sometimes even all committed, capital has been repaid to investors. In this scenario, however, the several years it could take for the fund's team to see any gains could lead to demotivation. An accepted compromise is that GPs take a lower percentage of early distributions until contributed capital is returned, either distributing the excess to LPs directly or putting it into an escrow account. Investors often require fund managers to escrow 20% to 30% of their carried-interest proceeds as a buffer against potential clawback liability.

A clawback is a liability triggered when, at the end of a fund's life, the LPs have received less than the sum of contributed capital and a certain amount of the fund's profits. Clawback is relevant in such situations as when early investments do well and repay more than the invested capital plus the preferred return, but later investments fail. A clawback ensures that managers will not receive a greater share of the fund's distributions than they are entitled to. Responsibility for payment of the clawback rests with the persons or entity that received the carried-interest distributions. In such a case, the GP is required to return some proceeds to make the investor whole.[10]

But clawback can also exist for LPs. In this case, it is triggered when, at the end of a fund's life, the GPs have received less than their share of the fund's profits. Such clawback is relevant in situations in which a portion of the committed capital has not been drawn, and LPs have received full repayment of their commitment plus hurdle before the GP has access to his or her carry.

7.7 CO-INVESTMENTS

Usually institutions investing in private equity seek intermediation, as few have the experience and especially the incentive structures that would allow them to directly invest in unquoted companies. To a lesser degree, institutions invest in companies directly or alongside PE funds in the form of co-investments. Some LPs generally look for such opportunities and even ask for co-investment rights, but co-investing is not an identical process to fund investing.

7.7.1 Direct Investments versus Co-Investments

Co-investing is often pitched to LPs as a way to reduce costs when they are waiting for opportunities to be presented to them, thus creating concerns regarding the quality of what they are being offered. In fact, a study put forward by Fang, Ivashina, and Lerner (2012) caught a lot of attention in the industry. These researchers find that direct investments have significantly outperformed standard fund investment benchmarks and notably co-investments:

- The evidence for underperformance of co-investments may be associated with the higher risk of such deals.[11]
- Importantly, the evidence found a sharp contrast between the performance of "solo" deals, in which institutional investors source direct investments independently from GPs, and that of co-investment deals.
- This points to an agency problem, because GPs may selectively offer deals to their LPs.

- Crucially, however, these results refer to an approach in which LPs took no active role in sourcing opportunities and only co-invested in deals offered to them, apparently based on deal size. (The deals were nearly five times larger than the funds' average deals.) Investments or products expected to perform far below expectations are often referred to as **lemons**.

That co-investments are viewed as being offered only by the funds to their LPs reflects a specific modus operandi that is not necessarily representative.[12] Practitioners point out that it is important that LPs generate their co-investment deal flows like any other direct investor would.[13] The PE market environment has a strong influence on the types of industries these co-investment opportunities stem from. Co-investment opportunities emerge particularly in periods when deal flow for larger investments is good but fundraising is difficult. In this situation and also when funds do not wish to be overexposed to one particularly large deal, GPs are likely to turn to their LPs and propose a co-investment.

LPs view co-investments as a way of enhancing their overall returns without having to pay a management fee or carried interest on their investment in the private company. LPs often refuse to pay, and fund managers charge low or no fees for access to co-investment opportunities and claim no carried interest but just transaction fees without setoff.[14] To be able to go this no-fee, no-carried-interest route, the LP needs to be an existing investor in the PE firm it invests alongside and either have a good relationship with it or be able to offer expertise in a particular field. However, the perception that all co-investment deals come without fees and carried interest is wrong. Indeed, some LPs find it wiser to **pay promote**, which encompasses the use of both carried interest and annual management fees on co-investments to keep the GP incentivized.

7.7.2 Expected Advantages of Co-Investing

In addition to avoiding the double layer of management fees, co-investing potentially offers the LP a number of advantages. Theoretically, it provides maximization of the fund investment's upside by increasing the exposure to the best-performing portfolio companies.

- Targeted investment tool: Commitments to PE funds are a relatively blunt allocation tool. Moreover, funds regularly experience so-called style drift over their lifetimes. Co-investments are a tool for building a targeted allocation to specific investments. If an LP follows the policy of co-investment only in portfolio companies that are already profitable or that would reach profitability soon, this could, for example, mitigate the exposure to cash-burning start-ups. Also, it may be difficult to find funds that specifically target, say, eco-innovation, but it is very likely that a number of funds have one or another company active in this sector in their portfolios. Co-investing allows the LP to tilt the portfolio balance in a desired direction or fill specific target allocations in the LP's portfolio.
- Better management of portfolio diversification: Co-investments can mitigate an institutional portfolio of funds' overdiversification or overexposure.

Co-investments provide flexibility to capitalize on industry-specific and country-specific opportunities as they arise. They may also be an answer to funds experiencing liquidity problems and can be a meaningful strategy with a first-time team, serving as a tool to assist fund managers in the execution of transactions. In the context of an overcommitment strategy, it allows an LP not to execute the option if the liquidity is insufficient to honor the resulting increase in capital calls. Moreover, it could be used for managing foreign exchange risks, as usual hedging products would be too uneconomical for PE funds, with their long lifetimes and volatile cash flows.

- Mitigating dilution: For investments in small funds, co-investments also improve downside protection, as they mitigate the dilution of a fund's shareholding when available resources are insufficient for follow-on investing. This can be attractive when funds lack sufficient capital for follow-on investment (particularly in a venture context), allowing them to boost their firepower on certain deals without resorting to collaboration with other GPs, which can potentially lead to conflicts and rivalry. However, if disputes on an investment arise, relationships with LPs and future fund raisings can be negatively affected.
- Dual level of review for investments: The institutional investor can leave the difficult work of deal sourcing and assessment to experienced industry experts. Even though LPs need to do some due diligence, they can save significant time, as much of the screening is done by fund managers, as is a large degree of the time-consuming task of monitoring.
- Improved monitoring: As an indirect benefit, a co-investment allows for improved monitoring of the funds and a further reduction of the information asymmetry between fund managers and their investors. It can help LPs better understand the investment process and environment, allowing for better fund selections and reinvestment decisions. The information co-investors receive is quite important because it gives an idea about how the fund managers operate and often goes beyond the standard reporting LPs usually receive.
- Establishing relationships to invitation-only funds: For LPs who face problems getting access to top funds, co-investments form an important part of a strategy that helps to land slots in a handful of sought-after funds. Moreover, an LP can benefit from the investment skills of such funds even if not yet invested and get access to transactions of top-tier PE firms.
- Reduction of the J-curve effect: Co-investments, like secondaries (see Chapter 9), are viewed as leading to a reduction of the J-curve effect and improved capital deployment and returns. Smoothing the cash-flow J-curve (that is, the concentrations in outflows due to capital calls during the early years of an investment program) may make sense in some situations but will be difficult to achieve, as there is usually too little and too inconsistent quality deal flow to have a meaningful impact on sizable portfolios.

Importantly, co-investing conforms to the important characteristics of real options and can maximize a fund investment's upside by increasing the exposure to the best-performing portfolio companies. However, it is definitely not a panacea for higher returns in the asset class and, in fact, is often seen as quite difficult to execute.

7.7.3 Expected Disadvantages of Co-Investing

Indeed, there are also disadvantages to take into consideration when engaging in co-investments:

- Unbalanced portfolios: Because an LP may build up additional exposure to certain companies, industries, or geographies, risks can increase very specifically with regard to individual portfolio companies. If few and too large co-investments are undertaken, there is also the danger of holding an overly concentrated portfolio of directly held investments.
- Increased fiduciary risk: For investment decisions taken outside a formal partnership structure, there is exposure to fiduciary risk. If things go wrong at the investment level, LPs may be exposed to legal liabilities and reputational risks.
- Conflicts of interest: There is a wide consensus that investors should focus solely on committing to the very best funds, and the potential to compensate through co-investment rights for higher perceived risks is limited. Moreover, co-investing can become a source for conflicts of interest. Co-investing LPs are likely to give preference to their specific investments even if this is detrimental to the entire fund's performance.
- Disagreements among LPs: Problems associated with failing co-investments can become complicated when one company with several co-investing LPs goes bankrupt, leading to disagreements between the parties involved. Also, fund managers may be inclined to spend more time on particular portfolio companies if they receive additional management fees or carried interest, or if the co-investing LP is of strategic importance for them.

For their risk-reward profile, co-investments are between a direct investment and a normal partnership. They are more risky than investing in a PE partnership, with a downside comparable to that of direct investments (while in a fund's portfolio, a failed investment can go under; also, co-investments result in more losers), and they require significant staff time and legal expenses.

7.7.4 Investment Process for Co-Investing

There is no set model for co-investments. The main difference between it and direct investing is how they are sourced. Opportunities are generally deals that a fund manager has prescreened, structured, and priced, and is expected to invest in.

- Portfolio strategies typically include a reserve of between 10% and 20% to be allocated to co-investments. Many investors do not appear to have a specific allocation and tend to do co-investments on an opportunistic basis, but a PE investment program requires a critical mass for being able to pursue co-investments.
- The fund generally organizes the co-investment for the LP, who needs a significant number of primary fund commitments to generate a meaningful co-investment deal flow. LPs need to express their interest in co-investments at the time of the fund commitment and reinforce this through regular follow-up calls and meetings.

- Importantly, the efficient execution of the process, be it turning a deal down quickly or working in parallel with and at the same pace as the GP, is critical to successful co-investing.
- Co-investing requires a different skill set than does fund investing, as LPs need to have a much greater insight into individual deals. If they are unable to form a view on individual investments, they will still be able to save management fees and carried interest but will not be able to significantly improve the returns of their portfolio of funds. In fact, LPs who strongly believe in their fund selection skills typically see little or no advantage to co-investing and often even argue against it.

For smaller LPs, the fact that an occasionally too high minimum amount is required for co-investments can cause problems. In such situations, the opportunity for co-investing may have to be passed up, especially if the amount required for co-investing is close to the amount the small institution originally committed to a fund. One way of overcoming this is to approach other investors and do the co-investment alongside them.

7.8 CONCLUSION

This chapter has provided an overview of the PE industry. Venture capital funds are high-risk, high-return investments in small growth companies. GPs of VC funds focus on the entrepreneur's business model and are often involved in the firm's board of directors and strategic planning process. Buyouts are lower-return, lower-risk investments focused on more mature companies. GPs investing in buyouts may seek to modify the firm's capital structure and implement operational improvements.

Funds of funds diversify over a number of funds. Although going this route adds a second layer of fees, funds of funds are a means to scale up as well as down and thus may be the best way for new and small investors to access the PE market.

GPs have a life cycle moving through three stages: (1) entry and establish, (2) build and harvest, and (3) decline, exit, or spinout. There can be high barriers to entry for new managers, but once established, GPs may be able to build long-term relationships with LPs who may invest in funds across several vintage years.

The LPA sets the economic incentives and penalties that make the agent (i.e., the fund manager) perform in line with the interests of the principals (i.e., the investors). As individuals, fund managers should do well when the funds they manage perform well, and much less well when the funds underperform.

Co-investing can be viewed as a way to overcome some of the restrictions of the limited partnership structure. However, co-investing is not a panacea and clearly is not for the risk-averse or inexperienced PE investor.

NOTES

1. The term *later stage* refers to the expansion, replacement, and buyout stages of investment.
2. Some in the investment industry use the term *private equity* to refer only to buyout investing, while others, as is done in this book, refer to both venture capital and buyout investing as private equity.

3. One could argue that there is implicit leverage through the intensive use of option-like mechanisms and due to the fact that there is constrained financing: start-ups are never fully financed, and seldom do funds have the financial resources to fund all their investments.
4. Jo (2002) analyzed 48 U.S.-based funds of funds launched between 1992 and 1999 (13 asset managers, 15 banks, and 20 independent funds). For asset managers, there was an average management fee of 0.85% and an average carried interest of 3.8% (only five of the 13 asset managers charged a carried interest). For investment and commercial banks, management fees were in the 0.88% to 1.25% range; 12 of the 15 banks charged a carried interest, with the average being 6.6% and the typical carried interest being just 5%. At the end of the 1990s, annual management fees were in the region of 0.8%, and carried interest was at 10%; five years later, the difficult market environment brought those down to 0.7% and 5%, respectively.
5. These principles were developed to encourage discussion between LPs and GPs regarding fund partnerships. They focus on alignment of interest, governance, and transparency.
6. See www.privateequityvaluation.com.
7. See www.peigg.org/valuations.html.
8. Associated with the Freedom of Information Act debate in the United States on transparency for public-institution limited partners, this approach is seen as a way out of the dilemma. It even goes so far as to oblige such limited partners to destroy the material provided by the fund manager as soon as it is practical for them to do so.
9. See Meek (2004): "Those that offer to put up the standard one per cent of the fund just because that is the standard, for example, are unlikely to convince investors that they truly believe in the investments they make. . . . Adveq's André Jaeggi agreed. 'The appropriate level of contribution depends on the fund manager's circumstances. If Kleiner Perkins partners agreed to put in ten per cent of the fund, that's a very significant amount, but it might not hurt the partners as much as a one per cent contribution made by newer players. . . . The general partner has to be at risk.' Or, in the words of another investor: 'I'm not interested in what the percentage is. I'm interested in how much it hurts the manager to lose the contribution.' "
10. Associated with this is joint and several liability: If a manager leaves the fund, and the fund ends up with a liability, this clause makes the remaining team members responsible for the departed person's share of the liability.
11. One explanation put forward for this result is that it reflects the fact that the sample consists of large and sophisticated investors. Small investors replicating a direct investment strategy may have different experiences.
12. For their study, Fang, Ivashina, and Lerner (2012) compiled a proprietary data set of direct investments and co-investments from seven large institutional investors.
13. See, for instance, Gallagher (2011).
14. In addition, funds of funds often do not charge carried interest or a special fee for co-investments, although there are exceptions.

REFERENCES

Fang, L., V. Ivashina, and J. Lerner. 2012. "The Disintermediation of Financial Markets: Direct Investing in Private Equity." Forthcoming in *Journal of Financial Economics*.

Fleischer, V. 2005. "The Missing Preferred Return." UCLA School of Law, Law & Economics Working Paper 465. February 22.

Gallagher, B. 2011. "Co-Investments in Funds of Funds and Separate Accounts." www.twinbridgecapital.com/pdf/m_Co-Investment.pdf.

Jo, H. 2002. "Perspectives and Problems of Private Equity Funds-of-Funds." Leavey School of Business and Administration.

Lerner, J. 2000. *Venture Capital and Private Equity: A Casebook*. New York: John Wiley & Sons.

Lerner, J., and A. Schoar. 2004. "The Illiquidity Puzzle: Theory and Evidence from Private Equity." *Journal of Financial Economics* 72 (1): 3–40.

Mathonet, P.-Y., and T. Meyer. 2007. *J-Curve Exposure: Managing a Portfolio of Venture Capital and Private Equity Funds*. Chichester, UK: John Wiley & Sons.

Maxwell, R. 2003. *To Disclose or Not to Disclose? That Is the Question*. London: AltAssets.

Meek, V. 2004. *Bottom of the Pile*. London: AltAssets.

Otterlei, J., and S. Barrington. 2003. "Alternative Assets—Private Equity Fund-of-Funds." Special Report, Piper Jaffray Private Capital.

Weidig, T., and, P.-Y. Mathonet. 2004. *The Risk Profiles of Private Equity*. Brussels: EVCA.

CHAPTER 8

Private Equity Benchmarking

Benchmarking aims to evaluate the performance of a specific entity by comparing it to a standard or a point of reference. Institutional investors in general, and institutional investors in private equity (PE) in particular, rely on benchmarks when making asset allocation decisions. PE managers rely on benchmarks as well.

A benchmark serves the following needs:

- As a component of the strategic asset allocation process, during which the risk-return properties of asset classes are evaluated and optimal allocation within a diversified portfolio is determined
- As a standard or point of reference against which the performance of a security or an investment manager can be assessed
- As a standard used by investment managers to compare their performance against in order to market their services to and communicate with current and potential investors
- As part of the process of providing the right incentives to managers, determining where and what kinds of improvements are needed in the investment process, and analyzing how competing managers achieve their high performance levels

The process of constructing and using a benchmark requires the following:

- Valuing a PE asset and measuring the performance of that asset
- Choosing the appropriate benchmark for a direct PE investment, a PE fund, or a fund of funds
- Comparing the performance of this PE asset with the benchmark

PE investments have key characteristics that distinguish this asset class from other asset classes and imply that both benchmarking and portfolio construction techniques will be different (see Chapter 9).

In the case of PE funds, benchmarking is normally performed using the past financial performance of a particular fund manager and, as such, forms part of the due diligence process. However, benchmarking can also cover current financial performance and is thus also used as a selection criterion (track-record analysis) in the monitoring process.

8.1 THE VALUATION OF PE ASSETS

The valuation of a PE investment can pose significant problems, given the often limited operating history of the investment, and is compounded in cases in which the company has yet to generate a profit. Private equity is mainly an **appraised asset class**, valued not by the consensus reflected in market prices being traded by many market players but by the professional opinions of a few experts. In this regard, PE shares some characteristics with other illiquid assets, such as private real estate funds. PE valuations are drawn up according to industry standards, such as the International Private Equity and Venture Capital Valuation Guidelines (IPEV Valuation Guidelines) introduced in 2005 and updated in 2015.[1]

Traditional valuation methods, such as discounted cash flow methodologies, can be applied to venture capital (VC) investments only by making numerous assumptions, often using rather unreliable information. The valuation of a VC investment is mainly based on the analysis of intangibles, such as patents or the founder's entrepreneurial skills, competence, and experience, as well as on the assessment of the expected market size for the portfolio company's products or the presumed exit value relative to existing comparable public companies. Thus, VC valuation is usually based not on cash flow or earnings but on multiples where comparable companies exist; where they do not, valuation becomes even more difficult.

There are relatively few investors and little or no consensus on valuation. A lack of third-party oversight, such as by debt providers, can make venture capital prone to losses from overvaluation. In addition, because the value placed on a young company cannot be verified except through future rounds of investment, it may take years to uncover overinflated and unsustainable valuations.

In buyout investments, valuation risk is more limited. To begin with, the valuation of portfolio companies is more straightforward, enabling one to choose from a rich toolbox of accepted instruments for quantitative analysis, such as discounted cash flows or multiples. The leverage required for transactions leads to scrutiny from a syndicate of commercial lenders and often due diligence by underwriters of a high-yield bond offering. The influence of these credit providers eliminates some of the potential risks inherent in the leverage. There will be restrictions on the amount of leverage lenders provide, which implicitly sets an upper boundary on the total valuation for the targeted business.

8.2 MEASURING PERFORMANCE OF PE FUNDS

Due to its unique characteristics, performance measurement in private equity is particularly challenging. The set of metrics commonly used to measure performance of other investments is flawed in the context of PE and does not give the right picture of a PE manager's performance.

8.2.1 Return Measures for PE Funds

The performance of most asset classes is measured as time-weighted returns, measuring the return that an investor achieves over some period of time, where each time interval carries the same weight independent of the amount of money invested. PE assets, in contrast, are described by their cash flows as investments and divestments that are spread out over time at the discretion of the manager. Therefore, a rate-of-return-based performance measurement along with a time-weighted evaluation is not appropriate. Instead, this requires measuring cash-weighted returns that correspond to the return obtained for a given cash flow pattern, where the same weight is given to each amount of money.

The main indicators for assessing PE performance are the internal rate of return (IRR) and the total value to paid-in ratio, which can be applied to portfolio company data and fund-level data alike. **Total value to paid-in (TVPI) ratio** is the ratio of the current value of remaining investments within a fund, plus the total value of all distributions to date, relative to the total amount of capital paid into the fund to date. Industry associations and CFA Institute deem the interim internal rate of return (IIRR), which is a cash-weighted IRR, and the since-inception IRR to be the most appropriate return measures for PE funds.[2]

The IRR is the discount rate that makes the present values (PVs) of the cash inflows (i.e., distributions) equal to the present values of the cash outflows (i.e., capital calls). Unlike the IRR method, where only the actual cash flows are employed to calculate the discount rate, the IIRR method includes the appraised residual value or (NAV) net asset value of the fund in the calculation process. The TVPI ratio, also called the "money multiple," is another way of measuring returns. The TVPI ratio's limitation is that it does not provide information on the investment period's duration. These measures can sometimes give very different pictures. Moreover, they allow comparisons within the PE universe, but not with standard asset classes such as public equity, where performance is measured as time-weighed rates of return. This makes direct comparisons between private equity and standard asset classes difficult, and global portfolio construction that combines them more complex.

To overcome the difficulty of direct comparison with publicly traded securities, the industry has developed a number of performance measures that take into account the performance of an external benchmark (e.g., the MSCI World Index). These are known as public market equivalent performance measures. Public market equivalent (PME) performance measures use external market indices to reflect the opportunity cost of capital. It is essentially assumed that the cash flows associated with a PE investment were invested in the external benchmark. Then the IRR or the multiple of the PE investment is calculated using the future value (FV) of cash flows as if they were invested in the public benchmark.

8.2.2 Internal Rate of Return

The IRR is the implied discount rate that makes the net present value of all cash flows equal to zero. The IRR is the most commonly used performance measure in private equity, but it has a number of well-known shortcomings, such as being very sensitive to both the timing and the size of cash flows. The IRR is particularly useful for measuring performance in situations in which investments and divestments are not regular, as when investing in PE funds. Mathematically, the IRR is found by solving the following equation:

$$\sum_{t=0}^{T} \frac{D_t}{(1+IRR_T)^t} - \sum_{t=0}^{T} \frac{C_t}{(1+IRR_T)^t} = 0 \tag{8.1}$$

where D_t is the fund distribution during the period t, C_t is the capital contribution or drawdown during the period t, and IRR_T is the investors' net internal rate of return at maturity T.

The interim IRR, which is the IRR of unliquidated funds, is a cash-flow-based return measure that considers the residual value or net asset value (NAV) of the partnership's holdings as a final cash inflow. IIRRs are only estimates rather than actual realized rates of return. Mathematically, the IIRR is found by solving the following equation:

$$\sum_{t=0}^{T} \frac{D_t}{(1 + IIRR_T)^t} - \sum_{t=0}^{T} \frac{C_t}{(1 + IIRR_T)^t} + \frac{NAV_T}{(1 + IIRR_T)^T} = 0 \tag{8.2}$$

where NAV_T is the latest net asset value of the fund at time T, and $IIRR_T$ is the investors' net interim internal rate of return at time T.

8.2.3 Modified Internal Rate of Return

Implicit in this performance measure is the assumption that distributions are reinvested at the IRR, which may be unrealistic and, if removed, may have a significant impact on the total return. To correct this weakness, the modified IRR (MIRR), which assumes a given reinvestment rate, is preferred by some investors. The modified IRR is the IRR calculated while taking into consideration the investor's cost of capital and reinvestment opportunities (see Exhibit 8.1). Mathematically, it is found by solving the following equation:

$$MIRR_T = \left(\frac{\text{FV of Positive Cash Flows Using the Reinvestment Rate}}{-\text{PV of Negative Cash Flows Using the Cost of Capital}} \right)^{1/T} - 1 \tag{8.3}$$

$$MIRR_T = \left(\frac{\sum_{t=0}^{T} D_t \times (1 + RR_T)^{T-1}}{\sum_{t=0}^{T} \frac{C_t}{(1 + C_0C)^t}} \right)^{1/T} - 1 \tag{8.4}$$

where RR_T is the expected reinvestment rate for the period until time T, C_0C_T is the investors' cost of capital for the period until time T, and $MIRR_T$ is the modified IRR for the period until time T. If the residual NAV of the fund is used in the numerator to calculate the future value of cash inflows as well as the fund's NAV, then the modified internal rate of return is obtained. Note the negative sign in the denominator of Equation 8.3, which highlights the fact that the absolute values of negative cash flows are used to calculate the present value.

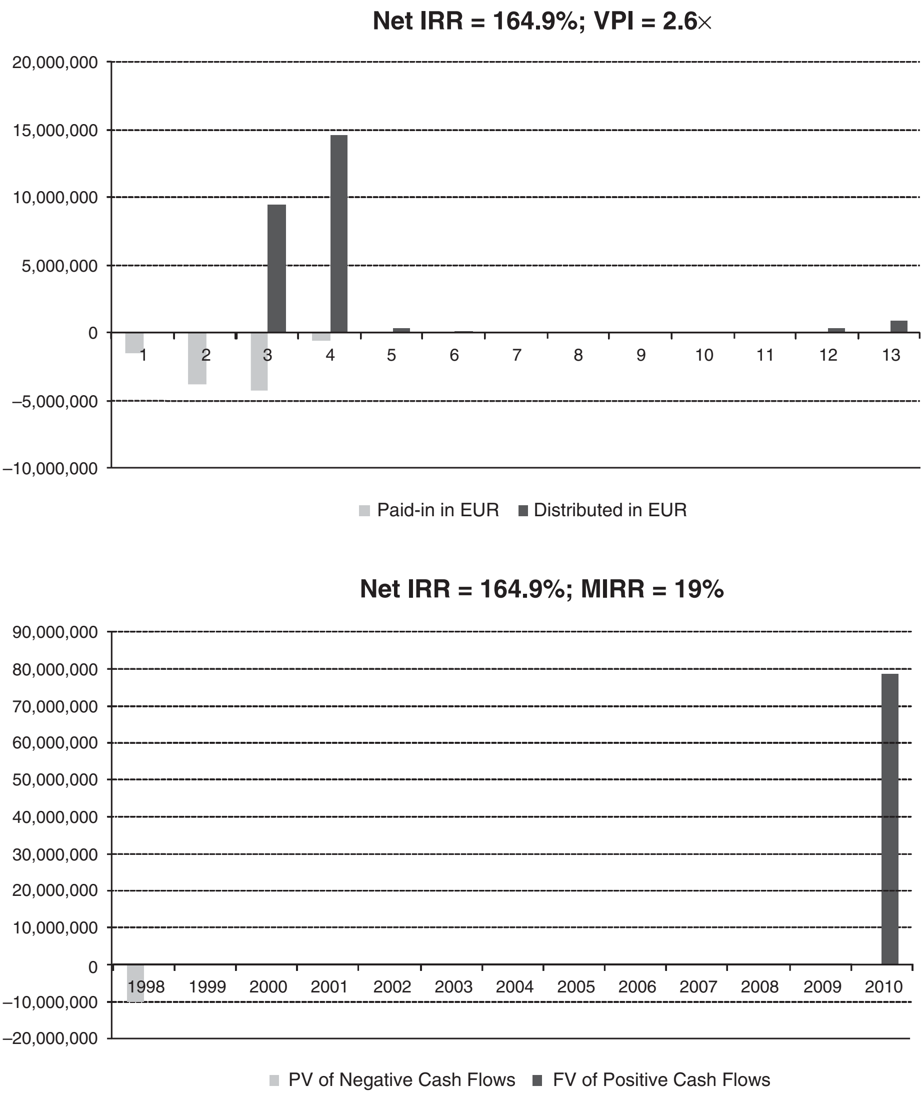


EXHIBIT 8.1 Modified IRR Example

8.2.4 Other Commonly Used Performance Measures

The IRR is not the only performance measure used by the industry. The following represent some of the other commonly used performance measures:

- The **distribution to paid-in (DPI) ratio,** or realized return, is the ratio of the cumulative distribution to investors to the total capital drawn from investors:

$$DPI_T = \frac{\sum_{t=0}^{T} D_t}{\sum_{t=0}^{T} C_t} \tag{8.5}$$

- The **residual value to paid-in (RVPI) ratio**, or unrealized return, at time T is the ratio of the total value of the unrealized investments at time T to the total capital drawn from investors during the previous time periods:

$$RVPI_T = \frac{NAV_T}{\sum_{t=0}^{T} C_t} \tag{8.6}$$

- The **TVPI ratio**, or total return, is a measure of the cumulative distribution to investors plus the total value of the unrealized investments relative to the total capital drawn from investors:

$$TVPI_T = \frac{\sum_{t=0}^{T} D_t + NAV_T}{\sum_{t=0}^{T} C_t} = DPI_T + RVPI_T \tag{8.7}$$

Note that these ratios are measures of net returns to invested capital and therefore do not take the time value of money into account, as distinct from the IRR.

It should be stressed that the PE industry does, in effect, attempt to appraise the NAV at the end of each quarter. IIRR, RVPI, and TVPI are computed based on these residual values. Their estimations are the most problematic components of return evaluation and are among the main reasons that quantitative benchmarking should be complemented by a qualitative analysis. DPI, which is measured using only distributed capital, is seen as the most reliable measure of returns, especially for more mature funds.

8.2.5 The J-Curve

One of the first concepts that investors in PE funds will encounter is the (in)famous J-curve, also referred to as the hockey stick (see Exhibit 8.2). The European Private Equity and Venture Capital Association (EVCA) defines the **J-curve** as the "curve generated by plotting the returns generated by a PE fund against time (from inception to termination)." The classic fund performance J-curve is caused mainly by the fact that valuation policies followed by the industry combined with the uncertainty inherent in PE investments prevent the revaluing of promising investments upward until quite late in a fund's lifetime, while fees, costs, and expenses are immediately deducted. As a result, PE funds tend to demonstrate an apparent decline in value during the early years of existence, the so-called valley of tears, before beginning to show the hoped-for positive returns in the later years of the fund's life. After about five years, the IIRR will give a reasonable indication of the final IRR at the fund's maturity. This period is generally shorter for buyout funds than it is for early-stage and expansion funds.

Some time ago, it was postulated that the introduction of the IPEV Valuation Guidelines would drive the J-curve to extinction, as a truly fair value for funds would eliminate the conservative bias caused by early expensing of costs and deferred

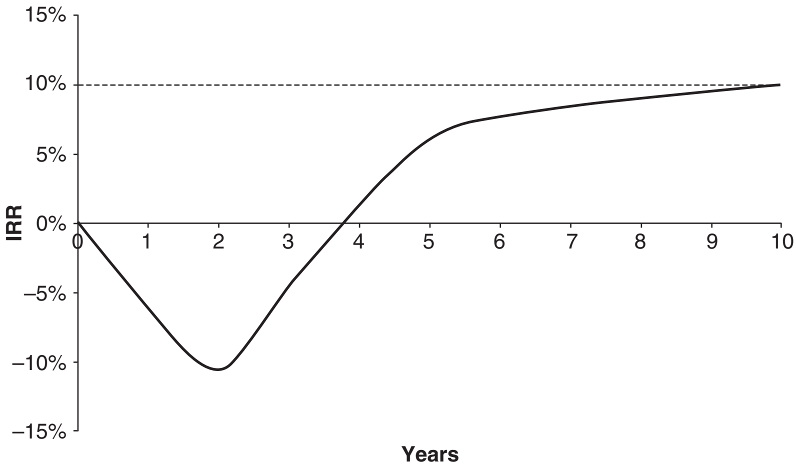


EXHIBIT 8.2 Fund Standard J-Curve

recognition of increases in the values of promising investments.[3] Instead, Mathonet and Monjanel (2006) found that the gap between the final IRR (or the expected IRR) and the IIRR narrowed in years 1 through 5, after which the IIRR became, on average, a reasonably reliable estimator of the final performance (see Exhibit 8.3).

But other J-curves can also be observed in PE funds: the cash flow J-curve and the net asset value J-curve. The NAV of a fund is calculated by adding the value of all the investments held in the fund and dividing by the number of outstanding shares of the fund. The **NAV J-curve** is a representation of the evolution of the NAV of a fund

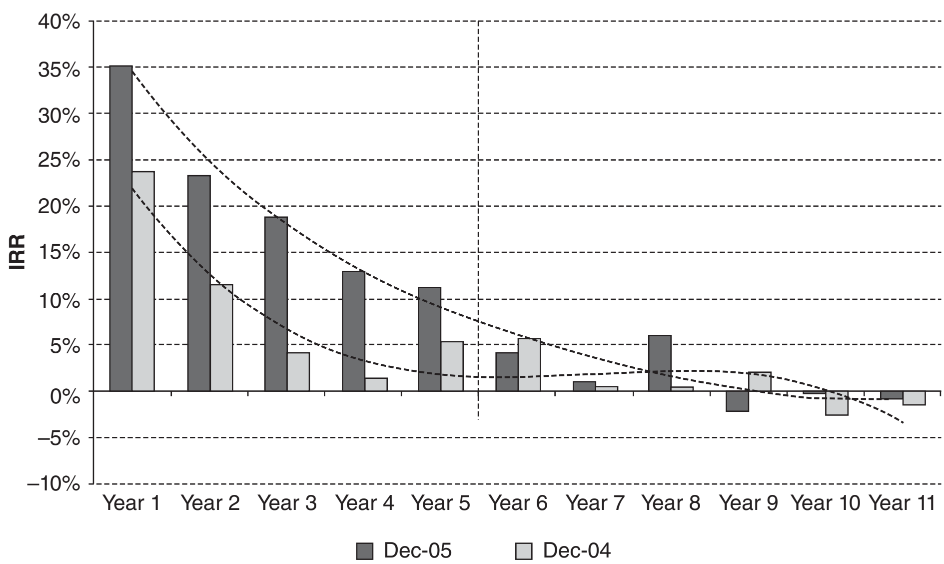


EXHIBIT 8.3 Old versus New J-Curve: Gap between the European Investment Fund (EIF) Portfolio's Final IRR Projections and Interim IRRs as of December 2005 versus December 2004
Source: Mathonet and Monjanel (2006) and European Investment Fund.

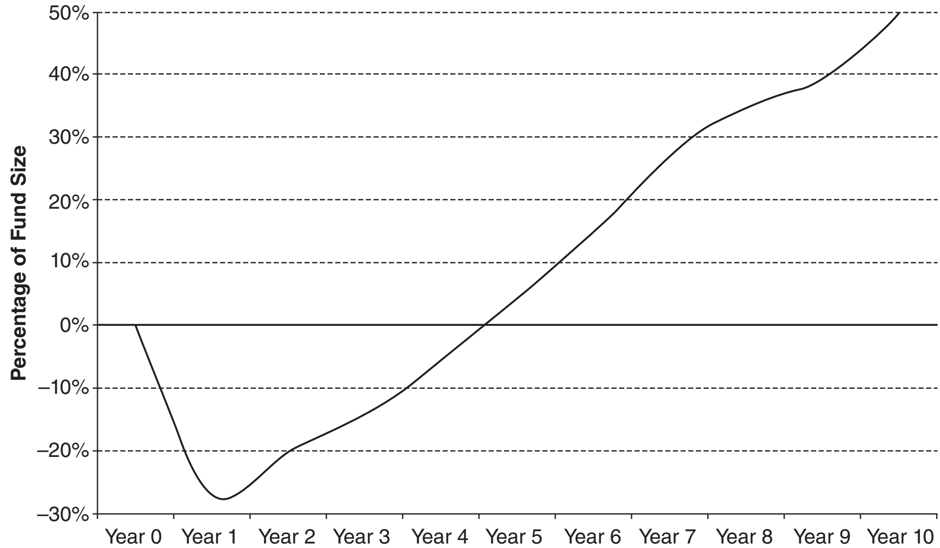


EXHIBIT 8.4 Cash Flow J-Curve

versus the net paid in (NPI), which first decreases during the early years of the fund's existence and then improves in its later years. The **cash flow J-curve** is a representation of the evolution of the net accumulated cash flows from the investors to the fund, which are increasingly negative during the early years of existence before making a U-turn and becoming positive in the later years of the fund's life. This is explained by the fact that in standard PE fund structures, commitments are drawn down as needed, or just in time, and when realizations are made after having successfully developed these newly founded companies, they are distributed as soon as is practical (see Exhibit 8.4).

8.3 BENCHMARK TYPES

Swamy, Zeltser, Kazemi, and Szado (2011) see two broad approaches to the construction of benchmarks, resulting in asset-based and **peer-group-based benchmarks**. An **asset-based benchmark** is constructed using public or private securities as its constituents. Equity, bond, commodity, and most real estate indices are of this type. For instance, the S&P 500 Index is constructed as a capitalization-weighted portfolio of the largest 500 of stocks traded in the United States. A peer-group-based benchmark is constructed using portfolios of investment managers as its constituents. For example, the CISDM Equity Long/Short Index is an equally weighted portfolio of hedge fund managers who follow this strategy.

Another categorization is according to whether returns are measured as absolute or relative:

- An **absolute return** is simply whatever an asset or a portfolio returned for a given period.
- A **relative return** is the difference between an absolute return and the returns on the asset or portfolio to which it is compared. Managers need to rely on relative benchmarks in order to assess how one asset compares to another.

Private equity is often perceived as an absolute return asset class, which is reflected in the carried interest that is the fund managers' main incentive and usually subject to a hurdle rate, or preferred return, so that it begins to accrue only once investors have received their capital back and a minimum pre-agreed-on rate of return (see Chapter 7).

8.4 ASSET-BASED BENCHMARKS

Many investors try to apply traditional quoted equity methodologies, in which risk and abnormal returns are estimated by time-series techniques, to private equity. Examples for asset-based benchmarks are listed PE indices and public equity indices. A **listed PE index** refers to an index whose components are share prices of publicly traded PE firms.

8.4.1 Listed Private Equity

In addition to data on listed private equity being available from stock exchanges, the listed private equity associations have developed their own indices, such as the LPX 50. Such indices track listed funds of funds, funds, and listed PE firms. They are based on up-to-date data, are investable, and feature daily trading; thus, they would normally satisfy the properties of an appropriate investment benchmark as defined in section 8.6.

However, these indices do not reflect the universe of assets in which institutional investors typically invest. For example, the S&P PE index contains Blackstone, an equity issue where revenues include the fees earned by the GP and not the return on the underlying investments. Moreover, listed private equity does not exhibit the same risk and return characteristics as nonlisted private equity. The statistical characteristics of listed PE indices, such as volatility and correlation to other assets, mainly depend on developments in the public markets (that is, short-term changes in supply and demand) rather than the intrinsic value of the underlying assets.

8.4.2 Public Equity Indices

As the typical PE assets are not investable, another approach is to take publicly traded securities that are believed to have the same risk exposures as a point of reference. Indeed, public equity indices are among the most common benchmarks applied in private equity (Nikulina, Calnan, and Disdale 2012). Such indices are believed to be exposed to the same market environments and have similar risk characteristics as private equity.

An often-used benchmark is a return target expressed as a premium over public equity (e.g., 300 to 500 basis points over public equity). This benchmark is a hybrid, being both a relative (public equity index) and an absolute (premium) benchmark. Its use is justified by the fact that the PE allocation is often obtained at the expense of public equity and thus can be considered a sort of perceived opportunity cost to private equity.

The concept of public market equivalent offers a simple solution to the benchmarking problem (see section 8.5). The **public market equivalent (PME)** is an index return measure that uses market indices to reflect the opportunity cost of capital.

The PME corresponds to the cash-weighted return that would have been achieved by investing in a stock market index at the same time that the PE fund makes a capital call and by selling index shares whenever the fund distributes capital back to investors. This investment strategy aims to replicate the irregular investment and divestment pattern of PE investing so as to allow for an approximate comparison of returns. Since PME is based on the same cash inflows and outflows of the PE fund, one may wonder why there would be any difference in returns. Actually, the difference comes from the net result of the investment and divestment processes in the index-tracking fund, which yields either a positive balance (public market shows higher performance) or a negative balance (public market shows lower performance).

8.4.2.1 Applying Public Equity Indices The foundation of various PME approaches is the use of a publicly traded index to calculate future values of cash flows as if they were invested in that index. For example, suppose the value of the publicly traded index at time t is given by I_t. Suppose D_t is the cash flow received at time t. If this cash flow were an investment in the index, its time T future value would be given by $D_t \times (I_T/I_t)$. The future value of all distributions and contributions can be calculated in the same manner:

$$FV(D) = FV(\text{Distributions}) = \sum_{t=0}^{T} \frac{D_t \times I_T}{I_T} \tag{8.8}$$

$$FV(C) = FV(\text{Contributions}) = \sum_{t=0}^{T} \frac{C_t \times I_T}{I_T} \tag{8.9}$$

One particular PME approach, which is known as **PME ratio**, uses the aforementioned future values according to the following equation:

$$\text{PME Ratio} = \frac{FV(D) + NAV}{FV(C)} \tag{8.10}$$

APPLICATION 8.4.2.1A

Suppose that the three-year cash flows of a PE investment are –100, +50, and +150, respectively. The NAV of the fund is 50 in year 3. Given that a public equity index had the values 120, 135, and 125 over the same three-year period, what are the $FV(D)$, $FV(C)$, and PME ratio over the three years?

The relevant future values of distributions $FV(D)$ and contributions $FV(C)$ are:

$$FV(D) = [50 \times (125/135)] + [150 \times (125/125)] = 196.3$$
$$FV(C) = 100(125/120) = 104.17$$

The PME ratio is therefore (196.3 + 50)/104.17 = 2.36.

Since the ratio is higher than 1.0, it is concluded that this private equity investment has outperformed the public market index by a substantial amount.

Depending on how the PME is calculated, there will be a number of issues and problems that must be carefully examined. These are discussed later in this chapter. Also, it is important to bear in mind that there are substantial differences between private equity investments and public equity investments (e.g., illiquidity); therefore, any direct comparison of a PE investment to public equity must be made with caution.

8.5 PEER GROUPS

The most intuitive approach is to compare a PE fund against a group of funds that have a similar risk profile (i.e., that have the same style or specialization)—that is, the fund's peer group. Peer groups, which are collections of similar funds used for benchmarking purposes, are seen as representative of the performance of the universe of PE assets available for selection, and they can be customized for various investment styles, such as U.S. buyouts, European venture capital, and emerging markets assumed to be subject to common return drivers.

8.5.1 Structure of PE Peer-Group Data

It is common practice in private equity to use a vintage, geographic, and stage- or focus-specific peer-group cohort, and to express a given fund's performance result in terms of the quartile it lies in within this cohort. A **peer-group cohort** refers to a group of private equity funds or investments that share some important characteristics. For example, one may consider European PE funds of a particular vintage to form a peer-group cohort. Other characteristics, such as industry focus, stage of investments, or even size, may be considered in creating such groups. A top performer in a dismal vintage year may barely return the invested funds, whereas in some spectacular vintage years, even fourth-quartile funds have returned double-digit returns. Exhibit 8.5 shows one example of individual fund benchmarking. The benchmarked fund starts as a fourth-quartile fund and moves after several quarters into the first-quartile area to peak with a 25% IIRR. It then goes down into the second quartile and ends its life at the boundary between the first and second quartiles.

A top-quartile fund is typically defined as a fund that belongs to the 25% best-performing funds in its peer group at the time of benchmarking. This leads some people to conclude that only 25% of funds may legitimately be qualified as top quartile. The fact is that many more funds in the market are being labeled and marketed as top quartile. One reason for this is that except for the 25% ratio itself, nothing else in this definition is carved in stone (see Chapter 9). For example, different fund managers or consultants may use different peer groups to calculate the relative performance of different funds. In addition, different periods may be used to compare the performance. Even if the same peer group and time period are used, different measures of performance (e.g., IRR, TVPI, or PME) may be used to compare performance. For example, while a fund may be ranked as a top-quartile fund using IRR, it may be ranked as a middle-quartile fund based on TVPI. It is thus good practice to benchmark a fund using a variety of measures but also to use judgment.

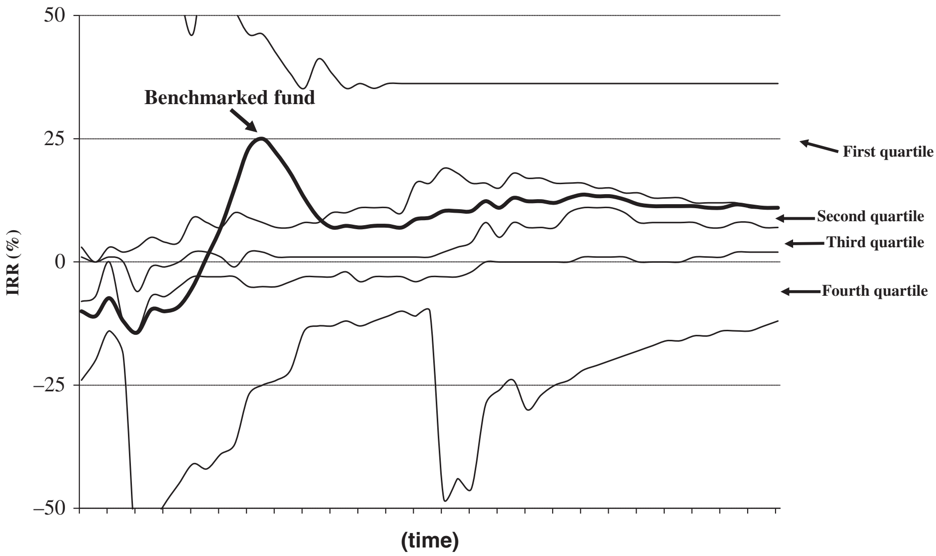


EXHIBIT 8.5 Individual Fund Benchmarking: Quartile Evolution

8.5.2 PE Peer-Group-Data Providers

Data of a PE fund's peers for benchmarking purposes are made available by various service providers, such as Cambridge Associates, Preqin, the Burgiss Group, Pitchbook Data, Bison, and previously Thomson VentureXpert.[4] All these databases suffer from shortcomings and are likely to have gaps:

- Databases are compiled mainly based on self-reported data provided by PE investors.
- Not all market participants are known, not all investors contribute data, and underperforming fund managers often withhold negative information and drop out. Therefore, it is difficult to assess how complete a picture of the market is being provided.
- Data quality is inconsistent and often very poor. In fact, only a few parties review and only infrequently work with these data, as funds and portfolio companies are usually not accessible for investment purposes.
- As some databases also cover funds of funds, their data could result in double counting.
- It is a general characteristic of an ever-changing alternative market space that innovations and emerging market spaces are initially not recognized as such, and the data relating to them are wrongly categorized or are simply ignored entirely.[5]

For these reasons, the service providers do not necessarily capture the same data, which can result in reported returns differing by as much as several percentage points for the same peer group.

Sometimes there are too few observations within one peer-group cohort to benchmark. Extending the peer-group cohort by closely comparable funds can be a simple way out. For example, if the number of 2005 European early-stage funds is not sufficient, the universe of all 2005 European VC funds or the universe of all 2004–6 European early-stage funds could potentially be used as alternatives.

8.5.3 PE Peer-Group Characteristics

Peer groups are not without problems. Due to the different fundraising cycles, the composition of the peer group of PE funds within vintages changes from year to year. Therefore, peer-group populations are not known in advance and are not unambiguous. Peer groups are typically not completely investable, as some funds may be closed to additional investors. Additionally, some funds' large minimum commitments prevent investing in the entire peer group.

Despite these issues, peer-group-based benchmarking is commonly used. Importantly for PE funds, peer-group-based benchmarking can overcome their various life-cycle-dependent distortions, notably the J-curve, as comparisons can be made between funds within the same stage of their development based on their interim returns. Geltner and Ling (2000) conclude that appraisal-based peer universe benchmarks are, in principle, valid and useful tools for investment performance evaluation.

8.5.4 Biases

Private equity is neither transparent nor efficient, which is one of the main arguments as to why it should outperform public markets. This lack of transparency is also reflected in the various biases these publicly available databases are subject to:

- Valuation bias: PE valuations are not to be confused with prices that can be observed in a marketplace. Since PE funds hold assets that are hard to value, their valuations by general partners give rise to the potential for these partners to manipulate reported NAVs. Managers may have an incentive to game valuations in the short run if returns on existing funds are used by investors to make decisions about commitments to subsequent funds managed by the same firm. As a result, the reported NAVs may be upwardly biased. Despite industry-led efforts to introduce stringent valuation standards, NAVs remain to a certain degree subjective and thus inconsistent.
- Stale pricing: In private equity, the reporting of financial statements is infrequent and can be delayed, often significantly. Therefore, the NAV may not reflect the most recent information on a fund.
- Instant history bias: Databases may contain performance figures that relate to a fund's incubation period. These returns tend to be inflated because it is at the discretion of the fund manager to report these figures.
- Selection bias: This bias stems from the fact that the benchmark is constructed using a sample that is different from the universe of all funds. For example, those managers that are closed to new investors may not report their performance to any public database. Despite the service providers' strong efforts over recent years to collect data from many PE funds, it is still difficult, if not impossible, for them to compile comprehensive data sets without any selection bias.
- Survivorship bias: This bias in particular can have a significant impact. The peer group overrepresents PE firms that have survived for an extended period of time, with funds that likely demonstrate superior performance, whereas firms that perform poorly tend to go out of business and therefore drop out of the peer universe. As a consequence, statistics will present an average historical performance that is likely to be upwardly biased.

However, most of the time practitioners can live with these biases and the benchmarks' other shortcomings or see them as inconsequential. According to Geltner and Ling (2000), it is not necessary to have a fixed and constant set of funds in the benchmark, and Swensen (2000) argues that survivorship bias may be less of a problem for long-term-oriented illiquid investments, such as private equity, as this population does not change rapidly. Indeed, managers of PE funds enter and exit the benchmark statistics with considerably less frequency than do their counterparts focusing on traditional marketable securities, since the limited partnership structure precludes an easy departure from the industry.

8.5.5 Controlling for Risk Differences

Managers can decide to be passive and structure a portfolio to closely mimic the market (or an index). By definition, following an active investment strategy implies that the risk level taken by the manager will be changing. If well-defined investable market indices or benchmarks do not exist, then each manager's investment strategy will need to have some degree of activeness. PE markets not only lack well-defined investable indices but also tend to be informationally inefficient. This means that not only do managers have to display some degree of activeness, but some active managers might be able to generate abnormal risk-adjusted returns on a consistent basis.

Usually there are two ways of controlling for risk differences: non-quantitatively, by constraining a manager to invest only in assets comparable to the peer group and then comparing the performance of the manager to the performance of that peer group; and quantitatively, by requiring the performance of the manager to meet certain quantitative standards and adjusting the ex post performance of the manager for risk. For instance, it may be required that the anticipated volatility, beta, or drawdown of the manager's performance should remain within a given range. Alternatively, the manager's ex post performance may be adjusted to compare its performance to a peer group or a public market index. In theory, the impact of risk can be controlled for by using risk-adjusted return measures in both the benchmark and the fund. However, since there are no widely accepted methodologies for obtaining risk-adjusted performance of PE funds, the non-quantitative method of constraining to a peer-group universe can, in practice, be used for controlling for differences in risk (Mathonet and Meyer 2007).

8.6 WHAT IS AN APPROPRIATE BENCHMARK?

As private equity is an appraised asset class, there are no universally accepted methods, and discussions continue on the validity of benchmarks.

8.6.1 Benchmark Properties

Bailey, Richards, and Tierney (1990) define the so-called **Bailey criteria** as a grouping of characteristics or properties that an investment benchmark should possess in order to be a useful gauge:

- UNAMBIGUOUS/KNOWABLE: The names and weights of entities that make up the benchmark should be clearly identifiable. PE benchmarks provide only aggregate data and do not give a complete representation of the available opportunity set.

- INVESTABLE: There should be an option to forgo active management and simply hold all assets that make up the benchmark. Beating an index, for illustration, would require holding assets that are not part of the benchmark or over- or underweighting assets relative to the benchmark. In private equity, one cannot choose, as in public markets, between being active and passive, as investors do not have access to the whole market. It is notably this fundamental difference that explains why PE portfolios exhibit widely diverging results.
- MEASURABLE: It is possible to frequently calculate the benchmark performance. Private equity data are insufficient for measuring risk and return characteristics accurately, as in the case of listed securities. Valuation guidelines, such as those put in place by various industry associations, define an appraisal policy for improving coherence and consistency, which makes comparisons between funds more meaningful. However, private equity remains subject to infrequent and subjective valuations, and interim performance figures are of limited value.
- SPECIFIED IN ADVANCE: The benchmark is constructed and mutually agreed on before the manager's evaluation. PE fundraisings are uncertain, so the weights and composition of any benchmark cannot be known in advance.
- APPROPRIATE: The benchmark is consistent with the manager's investment style. As private equity is essentially the search for overlooked opportunities, any benchmark will not fully represent the style, specialization, and expertise of the fund manager (such as in the case of emerging markets or VC investments in innovation, both of which, by definition, are not yet presented in the benchmark), and comparisons are problematic.
- REFLECTIVE OF CURRENT INVESTMENT OPINION: This requires understanding a benchmark enough to have opinions about whether to deviate from it. PE investors will not have full knowledge of the entities that make up the benchmark.
- OWNED: Investment managers have to agree with their sponsor that they are being measured against this benchmark and are accountable for the results.

PE benchmarks suffer deficiencies in nearly all of these dimensions. Tellingly, it is not market practice in private equity to base management incentives on performance relative to a benchmark; in addition, PE benchmarks are far less integral to the typical investment process than benchmarks for liquid assets are.

8.6.2 Benchmark Selection

Two primary factors need to be considered when selecting an appropriate benchmark: How is exposure to private equity sought, and whose success is measured? According to Nikulina, Calnan, and Disdale (2012), this depends on how private equity is invested:

- DIRECTLY INVESTING IN PORTFOLIO COMPANIES: Doing *solo deals* (i.e., when an internal PE investment team directly invests in portfolio companies) is becoming increasingly common among larger institutions. In this case, the benchmarking aims to assess an internal PE investment team's ability to access, select, and invest in high-quality portfolio companies that are within the profile of the stated mandate, and how much value the team can add to them. A public market benchmark plus a premium is meaningful, but preference should be given to the long-term equity market return. This would be under the condition that the investor is not subject to the same need to generate liquidity as is a PE manager acting on behalf of third-party investors.

- Investing through funds: When investing in private equity using funds as intermediaries, benchmarks relate to the fund managers' ability to access, select, and invest in high-quality portfolio companies compared with their peers and to add value to them. Here, peer groups as well as indices for benchmarking can be used. Fund managers should be assessed on whether they are able to outperform their peers. A public equity benchmark (as measured, for instance, through the PME) should, however, be introduced only to complement the peer-group comparison once the fund is substantially invested.
- Investing through a fund of funds: The alternative to an internal PE investment team is outsourcing the investment into funds to a fund-of-funds specialist. The fund of funds should be benchmarked to its opportunity set (that is, the PE fund population available during the years the manager has to commit its capital). This comparison needs to take the fund of funds' additional fee layer into account.

PE performance needs to beat the alternative use of capital in the context of a strategic asset allocation. Many investors consider private equity to be a return enhancer, as opposed to a diversifier to listed equities, and therefore demand a long-term return in excess of public equities. Nikulina, Calnan, and Disdale (2012) point out that public market performance serves as a measure of the opportunity cost, as private equity is often funded out of public equities. Here, public equity defines the cost of capital for PE investing. From this perspective, a suitable benchmark would be a public equity index or a long-term approximation of equity market returns, plus a premium required to compensate for private equity's illiquidity.

8.7 EXAMPLE FOR BENCHMARKING PE FUNDS

Let us take a look at Exhibit 8.6 to illustrate the use of the concepts just presented. Suppose that we have the listed values for distributions, contributions, and NAVs for two PE European funds (named PE Fund 1 and PE Fund 2) that belong to the vintage-year 2000 buyout funds.

Positive numbers correspond to years in which investors received net distributions, negative numbers correspond to years in which investors made net contributions, and the figures for 2006 correspond to the NAVs of each of the two funds at

EXHIBIT 8.6 Simulated Distribution of Two Funds (€ million), 2006 values are NAVs

	2000	2001	2002	2003	2004	2005	2006
PE Fund 1	–200	–800	200	–2,000	–600	2,000	3,500
PE Fund 2	–1,500	–1,500	–800	–200	500	1,500	5,000

the end of that year. To benchmark these funds, perform the steps illustrated in the next four sections.

8.7.1 Apply IIRR to Examples of PE Fund Returns

Recall that the interim IRR is defined as the discount rate that makes the present value of the distributions, the contributions, and the NAV equal to zero. More formally, the IIRR is found by solving the following equation:

$$\sum_{t=0}^{T} \frac{D_t}{(1+IIRR_T)^t} - \sum_{t=0}^{T} \frac{C_t}{(1+IIRR_T)^t} + \frac{NAV_T}{(1+IIRR_T)^T} = 0 \tag{8.11}$$

Solving this equation using a financial calculator or Excel (function IRR or XIRR), we obtain an IIRR of 12.53% for PE Fund 2 in this example. Using a standard financial calculator, the cash flows are entered for each time period, and the calculator performs the trial-and-error search when the IRR key is selected and the compute key is pressed.

APPLICATION 8.7.1A

What is the IIRR of PE Fund 1?

As just noted, a solution may be found using a financial calculator by entering the cash flows (and NAV if applicable) and computing IRR following the same procedure as for PE Fund 2. Computation for PE Fund 1 indicates an IIRR of 16.53%, which is more profitable than PE Fund 2 by exactly 4%. However, it can be essential to take into account that the cash flows of PE Fund 1 from left to right have three sign changes. Therefore, there can be up to three different IIRRs that solve for Equation 8.11 in the case of PE Fund 1. In cases of IIRR computations with multiple sign changes, none of the solutions can be trusted.

Note that one would need to compare these IIRRs to the discount rates or the required rates of return applicable to each PE fund in order to determine whether these returns were greater than the required minimum returns. (Further discussion on the discount rates applicable in private equity appears in Chapter 13.)

8.7.2 Apply the TVPI, the DPI, and the RVPI to Examples of PE Funds

TVPI: In the case of PE Fund 1, the TVPI is found by inserting the cash flows into Equation 8.7:

$$TVPI_T = \frac{200 + 2{,}000 + 3{,}500}{200 + 800 + 2{,}000 + 600} = 1.58 \tag{8.12}$$

The ratio indicates that 58% more cash is received (or indicated by an NAV) than is paid out.

APPLICATION 8.7.2A

What is the TVPI for PE Fund 2?

Inserting the cash flows and NAV from the previous exhibit into Equation 8.7 indicates that in the case of PE Fund 2, the TVPI is 1.75.

Thus, PE Fund 2 has a higher ratio of total distributions and NAV to total contributions between 2000 and 2006 than does PE Fund 1. As mentioned previously, this measure does not take into account the time value of money. Also, note that even though the drawdowns or paid-in capital are negative in the exhibit (given that they represent a use of cash to PE funds), their values are expressed in positive numbers in the denominator of the equation. This convention is followed because it generates a more meaningful sign (i.e., a positive value) for the TVPI index, which is more easily interpreted than benefit-to-cost ratios are usually expressed and interpreted.

The same procedure is followed when calculating the total value of drawdowns in the case of the next two indices (DPI and RVPI). For DPI, the distributions are inserted into Equation 8.5. In the case of PE Fund 1, the DPI is determined as follows:

$$DPI_T = \frac{200 + 2{,}000}{200 + 800 + 2{,}000 + 600} = 0.61 \tag{8.13}$$

APPLICATION 8.7.2B

What is the DPI for PE Fund 2?

Inserting the distributions and contributions from the previous exhibit into Equation 8.5 indicates that in the case of PE Fund 2, the DPI is 0.50.

DPI analysis indicates that PE Fund 1 has a higher ratio of total distributions to total commitments between 2000 and 2006 than does PE Fund 2. As mentioned previously, these measures do not take into account the time value of money (except partially through the use of an NAV in the RVPI and TVPI).

Finally, use Equation 8.6 when calculating the residual value to paid-in (RVPI) ratio:

RVPI: For PE Fund 1, the RVPI is:

$$RVPI_T = \frac{3{,}500}{(200 + 800 + 2{,}000 + 600)} = 0.97 \tag{8.14}$$

APPLICATION 8.7.2C

What is the RVPI for PE Fund 2?

Inserting the NAV and contributions from the previous exhibit into Equation 8.6 in the case of PE Fund 2, the formula gives us an RVPI of 1.25.

It can be seen that PE Fund 2 has a higher ratio of NAV to total contributions than does PE Fund 1. Again, note that this measure does not consider the time value of money. However, these three ratios can be useful general indicators of the performance of each of the PE funds.

8.7.3 Apply a Classic Benchmark Analysis to Examples of PE Funds

A classic benchmark analysis can be performed based on the following information collected for 31 European PE funds categorized as vintage-year 2000 buyout funds, from inception to December 31, 2006:

- The maximum return (measured using the IIRR) registered by a PE fund was 34.80%.
- The highest quartile of PE funds had a return of 13.20% or more.
- The median return was 6.50%.
- The lowest-quartile funds had returns of 0% or less.
- The minimum return was −9.50%.

Based on the information provided, a table can be constructed in Exhibit 8.7 to help visualize the performance of PE Funds 1 and 2 using a classic benchmark analysis.

It can be seen that PE Fund 1 had an excellent return when compared to its peers, as its IIRR was located in the top quartile of returns corresponding to the 31 European PE funds used in the sample. In the case of PE Fund 2, the observed return was less impressive, although its IIRR was still above the median PE fund return of the sample.

8.7.4 Apply a Public Market Equivalent Analysis to Examples of PE Funds

Compare the returns offered by the two PE funds to those of the public index, calculating the gap between the IIRR of each PE fund and the public market equivalent. The value of the public index during the 2000–6 period is shown in Exhibit 8.8.

This form of PME uses the IRR to calculate performance relative to a public index. Recall the formulas for $FV(D)$ and $FV(C)$ from Equations (8.8) and (8.9). *Note:* The terminal values do not enter the calculations $FV(D)$ and $FV(C)$. Then define NAV_{PME} as the difference between these two figures. Next, calculate the IIRR of the PE funds using their cash flows and the value of NAV_{PME} as its interim NAV,

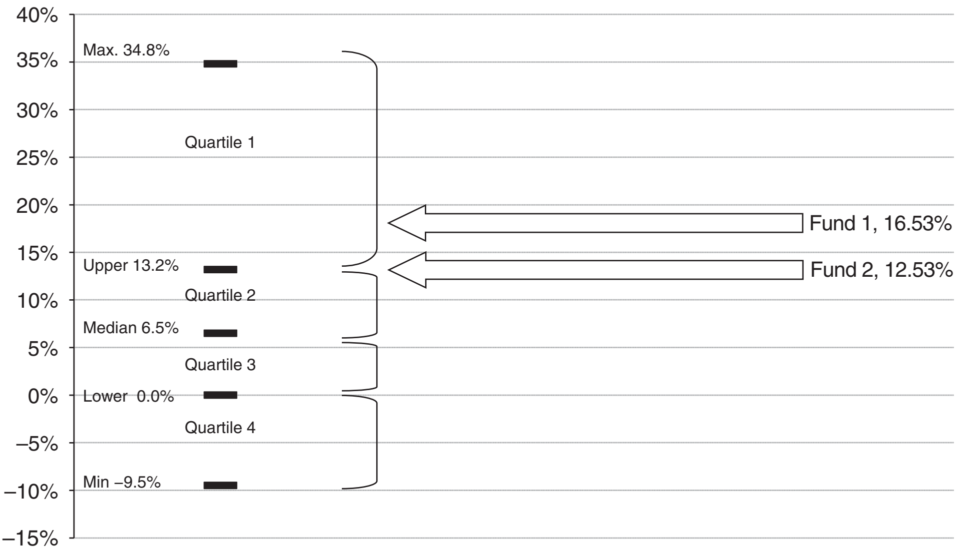


EXHIBIT 8.7 Performance Benchmarking

while ignoring the actual NAV of the fund. Finally, this IIRR is compared to the actual IIRR of the fund. In this example, the NAV_{PME} of PE Fund 1 in Exhibit 8.9 is €2,419, and that of PE Fund 2 is €2,471.

The IRR of PE Fund 1 is 16.5%. The annual growth in the public index from 2000 to 2006 is −1.1%. As stated previously, PE return measures, which are cash-flow weighted, cannot be compared directly to publicly quoted asset return measures, which are time weighted. To make this comparison possible, we use the public market equivalent (PME) methodology, which allows the practitioner to convert the time-weighted measures used for publicly quoted assets into a cash-flow-weighted methodology, which can then be used to benchmark a PE fund. An analyst should also consider the dividends paid by the stocks tracked by the public market index. In order to more accurately compare private equity and public equity, a total return index, which includes dividends, should be used for each.

In our simplified examples (we assumed only one cash flow per year), PE Fund 1 had a series of cash flows between 2000 and 2006 and a terminal interim value (i.e., the NAV of the fund) of €3,500 in 2006, leading to an IIRR of 16.5%. During the same period, a public market index (which is representative of the strategy followed by this fund) had a time-weighted return of –1.1% per annum, suggesting a significant underperformance versus the fund. The PME index is obtained by buying index units for each drawdown and selling index units for each distribution for an equal cash-flow amount. For example, in 2000, €200 was drawn, which if invested

EXHIBIT 8.8 Public Index Value, 2000–6

	2000	2001	2002	2003	2004	2005	2005
Public index	5,926	4,625	3,064	3,558	3,821	4,715	5,542

			2000	2001	2002	2003	2004	2005	2006	IIRR
PE Fund 1	Drawdowns	(1)	–200	–800		–2,000	–600			
	Repayments	(2)			200			2,000		
	Terminal NAV	(3)							3,500	
	"Cashflows"	(4) = (1) + (2) + (3)	–200	–800	200	–2,000	–600	2,000	3,500	**16.5%**
Public Market Equivalent	Index Value	(5)	5,926	4,625	3,064	3,558	3,821	4,715	5,542	–1.1%
	Acquired units	(6) = –(1) / (5)	0.03	0.17		0.56	0.16		X	
	Sold units	(7) = –(2) / (5)			–0.07			–0.42		
	Cumulated units	$(8)_t = (6)_t + (7)_t + (8)_{t-1}$	0.03	0.21	0.14	0.70	0.86	0.44	0.44	
	Investment	(1)	–200	–800		–2,000	–600		=	
	Desinvestment	(2)			200			2,000		
	Terminal Portfolio value	(9) = (8) × (5)							2,419	
	"Cashflows"	(10) = (1) + (2) + (9)	–200	–800	200	–2,000	–600	2,000	2,419	**9.1%**
		(4) – (10)			**Over (+) or under (–) performance**					**7.4%**

			2000	2001	2002	2003	2004	2005	2006	IIRR
PE Fund 2	Drawdowns	(1)	–1,500	–1,500	–800	–200				
	Repayments	(2)					500	1,500		
	Terminal NAV	(3)							5,000	
	"Cashflows"	(4) = (1) + (2) + (3)	–1,500	–1,500	–800	–200	500	1,500	5,000	**12.5%**
Public Market Equivalent	Index Value	(5)	5,926	4,625	3,064	3,558	3,821	4,715	5,542	–1.1%
	Acquired units	(6) = –(1) / (5)	0.25	0.32	0.26	0.06			X	
	Sold units	(7) = –(2) / (5)					–0.13	–0.32		
	Cumulated units	$(8)_t = (6)_t + (7)_t + (8)_{t-1}$	0.25	0.58	0.84	0.89	0.76	0.45	0.45	
	Investment	(1)	–1,500	–1,500	–800	–200			=	
	Desinvestment	(2)					500	1,500		
	Terminal Portfolio value	(9) = (8) x (5)							2,471	
	"Cashflows"	(10) = (1) + (2) + (9)	–1,500	–1,500	–800	–200	500	1,500	2,471	**2.5%**
		(4) – (10)			**Over (+) or under (–) performance**					**10.0%**

EXHIBIT 8.9 Performance Benchmarking Using Public Equity

in the public market index would have allowed the purchase of 0.03 index units. In 2002, the fund distributed €200, which if divested from the public market index would have required selling 0.07 index units. At the end of 2006, the terminal value of the PME is obtained by calculating the remaining index units and valuing them at the terminal index value.

In our example, we have 0.44 (rounded from 0.4365) remaining index units, which at €5,542 per unit are valued at €2,419. This contrasts with the fund investor, who holds €3,500 at the end of the period. Based on this, we can then calculate an IIRR for the PME (note that except for the terminal value, all the cash flows are identical to those of the fund), which can then be compared to the IIRR of the fund. The IIRR for the public market equivalent is 9.1%. We find that the gap, or excess IIRR, for the return of PE Fund 1, 16.5%, relative to the PME, is 7.4%. For PE Fund 2, the IIRR for the public market equivalent is 2.5%, underperforming the fund's IIRR of 12.5% by 10%.

We can use the same information to calculate the PME ratios of the two funds. PE Funds 1 and 2 have an $FV(D)$ of €5,131 and €4,959, respectively, and an $FV(C)$ of €7,438 and €6,568, respectively. Given their interim NAVs, the PME ratios of the two funds are 1.16 and 1.52. This indicates that the two funds outperformed the public index and that PE Fund 2 performed better than PE Fund 1 using the PME ratio.

8.8 PORTFOLIO OF PE FUNDS

To benchmark a portfolio of PE funds, the portfolio needs to be compared against a similar portfolio of PE funds. There are, however, two problems with this comparison: (1) publicly available database providers report too few funds of funds to make a comparison meaningful, and (2) these funds of funds implement various investment strategies, have different portfolio compositions, and, most problematic, usually have a different vintage-year structure. To circumvent these problems, synthetic portfolios can be generated with the same allocation to the various asset subclasses (e.g., vintage year, stage, and geographies) as the one to be benchmarked. Such benchmarking allows one to evaluate the portfolio manager's selection skills (i.e., how good she was at selecting the best fund managers within the defined allocations). If the portfolio is composed of 40% buyouts and 60% VC, the synthetic portfolio would need to have this same 40/60 split.

As the performance of a portfolio is the aggregation of individual funds' performance, the benchmarking of a portfolio is simply the extension of the benchmarking of an individual fund. When benchmarking, it is important to use the same aggregation method for both the portfolio and the benchmark.

8.8.1 Performance Measures

Because a PE portfolio is an aggregation of funds, its performance measures are simply the aggregation of the measures used for the individual funds (IIRR, TVPI, DPI, or RVPI), and can be calculated based on one of the following methods:

- Simple average: the arithmetic mean of the PE funds' performance measures:

$$IIRR_{P,T} = \frac{1}{N}\sum_{t=1}^{N} IIRR_{i,T} \tag{8.15}$$

Here, $IIRR_{P,T}$ is the IIRR of portfolio P at the end of period T, $IIRR_{i,T}$ is the IIRR of fund i at the end of time period T, and N is the number of funds in the portfolio. Note that this average IIRR may give misleading signals about the performance of the portfolio. For example, the simple average would be different from the IIRR calculated using the average cash flows. Therefore, the simple average IIRR has to be used with caution.

APPLICATION 8.8.1A

Consider Fund A with an investment of \$100 in year 0 and an NAV of \$110 in year 1, and Fund B with an investment of \$200 in year 0 and an NAV of \$288 in year 2. What is the simple average of the IIRRs for a PE portfolio composed of these two funds?

The IIRR of Fund A is 10%, and the IIRR of Fund B is 20%, so the simple average of the IIRR is 15%.

- MEDIAN: The value appearing halfway in a table ranking the performance of each fund held in the portfolio
- COMMITMENT WEIGHTED: The commitment-weighted average of the funds' performance measures:

$$IIRR_{P,T} = \left(\frac{1}{\sum_{i=1}^{N} CC_i} \right) \sum_{i=1}^{N} CC_i \times IIRR_{i,T} \tag{8.16}$$

Here, CC_i is the commitment made to fund i. In this case, the IIRR of each investment is weighted by the relative size of the commitment of each investment. Note that this calculation works for either IIRR or IRR.

The **commitment-weighted IRR** is an average calculated by weighting the rates of return by commitment. The commitment-weighted IRR does not capture the funds' actual investment. It is useful for measuring the LP's skill in selecting and allocating the right amount to funds.

APPLICATION 8.8.1B

Consider Fund A with an investment (and commitment) of \$100 in year 0 and an NAV in year 1 of \$110, and Fund B with an investment (and commitment) of \$200 in year 0 and an NAV of \$288 in year 2. What is the commitment-weighted average of the IIRRs for a PE portfolio composed of these two funds?

The IIRR of Fund A is 10%, and the IIRR of Fund B is 20%. The commitment-weighted average of the IIRRs is 16.67%, found as [(100/300) × 10%] + [(200/300) × 20%].

- Pooled: Portfolio performance obtained by combining all individual funds' cash flows and residual values, as if they were from one single fund, and solving the equation for $IIRR_{P,T}$:

$$\sum_{t=0}^{T}\sum_{i=1}^{N}\frac{CF_{i,t}}{(1+IIRR_{P,T})^{t}}+\sum_{i=1}^{N}\frac{NAV_{i,T}}{(1+IIRR_{P,T})^{T}}=0 \tag{8.17}$$

Here, $CF_{i,t}$ is the net cash flow during period t between the fund i and the investor, T is the number of periods, $NAV_{i,T}$ is the latest NAV of the fund i, $IIRR_{P,T}$ is the IIRR of the portfolio P at the end of time period T, and N is the number of funds in the portfolio.

The **pooled IRR** is a measure that attempts to capture investment timing and scale, and is calculated by treating all funds as if they were one composite fund. This composite fund's cash-flow series is then used to calculate the pooled IRR. The advantage of this measure is that it takes the scale and timing of cash flows into account and therefore also reflects the LP's skills to time the market. The disadvantage is that larger cash flows will be given more weight, so in a composite portfolio of small early-stage funds and large later-stage or buyout funds, the larger funds will distort the picture.

APPLICATION 8.8.1C

Consider Fund A with an investment of \$100 in year 0 and an NAV of \$110 in year 1, and Fund B with an investment of \$200 in year 0 and an NAV of \$288 in year 2. What is the pooled IIRR for a PE portfolio composed of these two funds?

The cash flows are −300 in year 0, +110 in year 1, and +288 in year 2. Using a calculator to perform the search, the pooled IIRR is 18.01%.

Arguably, the pooled measure gives the true financial return of the portfolio. However, it may also make sense to use the others, depending on what one wishes to measure. For example, the simple average can be a good indicator of the selection skills, while the commitment-weighted average can be useful in assessing the added value resulting from the decision of what size commitment to make to each specific fund.

Finally, in some cases, the IRR may fail to properly assess long-term performance. This stems from the fact that when IRR is positive, large inflows in the later years will not boost the return, but when IRR is negative, they will in absolute terms. This further explains why it is important to also assess the performance with, for example, the multiple. An alternative for this problem is to use the time-zero IRR, which is a pooled IRR calculated with the assumption that all portfolio investments begin on the same date. This is used to prevent the order of investments from affecting

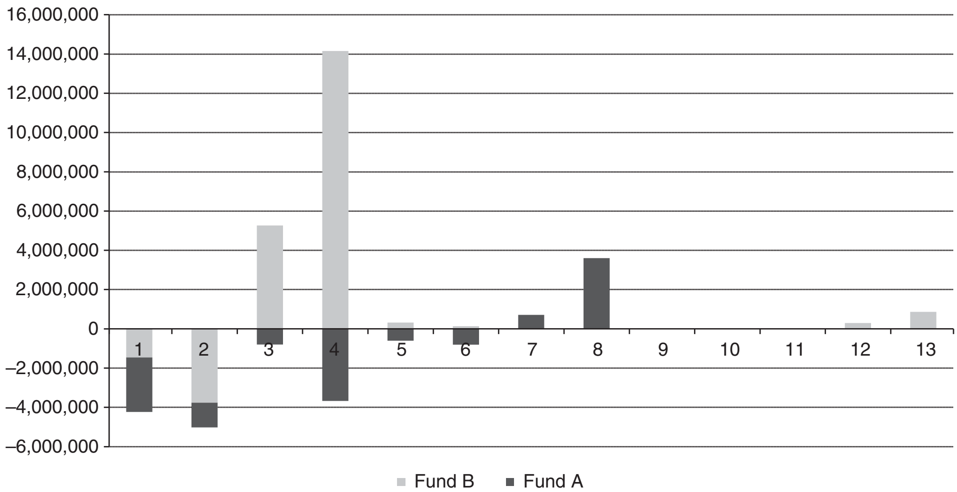


EXHIBIT 8.10 Time-Zero Net IRR: Two-Fund Portfolio Example (Fund A and Fund B)

the portfolio IRR (see Exhibit 8.10). The IRR methodology can lead to multiple or incorrect answers when cash flows change signs at various times. The problem can be compounded when using the pooled approach, as multiple funds over multiple time periods are more likely to cause misleading IRR calculations.

Performance for portfolios of funds is measured, for instance, as a pooled IRR or as a commitment-weighted IRR. Exhibit 8.11 illustrates all of these aggregated performance measures for a three-fund portfolio.

EXHIBIT 8.11 Aggregated Performance Measures for a Three-Fund Portfolio

	Commitment	2008	2009	2011	2012	2013	2014	2015	IRR
Fund 1	200	–50	–150	50	150	250			36%
Fund 2	300		–200	–100	100	200	250		23%
Fund 3	500			–300	–200	100	100	250	–4%
Sum	1,000	–50	–350	–350	50	550	350	250	16%
		0	1	2	3	4			
Time Zero		–550	–450	250	450	750			12%

Summary

Simple Average	19%
Median	23%
Commitment-Weighted	12%
Pooled	16%
Time Zero Pooled	14%

8.8.2 Determining Portfolio Benchmarks

In addition to the issue of how to compute the IIRR of an investor's portfolio of PE funds, there is the issue of how the IIRRs in a peer group are aggregated to determine the IIRR to be used as benchmark. Consider the case of commitment-weighted IIRRs. To compare apples to apples, one must compare the commitment-weighted portfolio performance to that of the commitment-weighted benchmark.

The portfolio benchmark can be constructed using the commitment-weighted average of the benchmark for each individual fund constituting the peer-group cohorts (e.g., the same vintage, geographic, and stage focus):

$$BM_{P,T} = \left(\frac{1}{\sum_{i=1}^{N} CC_i} \right) \sum_{i=1}^{N} CC_i \times BM_{i,T} \tag{8.18}$$

where $BM_{P,T}$ is the portfolio benchmark at the end of time period T, CC_i is the commitment to the fund i, N is the number of funds in the portfolio, and $BM_{i,T}$ is the benchmark of fund i at the end of time period T.

If all funds invested all their capital during the first takedown, the pooled return and commitment-weighted return would be identical. Additional confusion is caused by the question of whether to consider the impact of the funds' undrawn commitments. Neither the pooled nor the commitment-weighted IRR reflects how the LP manages the capital that is dedicated to investments in private equity but not called by the funds.

8.8.3 Monte Carlo Simulation

Monte Carlo simulation[6] is one technique that can be used to generate portfolios similar to the one being benchmarked by drawing, at each simulation run, the same number of funds that are in the portfolio out of all the relevant peer-group cohorts, and then weighting the performances by the commitment size of the funds in the portfolio. For example (ignoring for simplification reasons the vintage year and the geographical risk dimensions), for a portfolio composed of eight early-stage funds and five later-stage funds, the simulation will draw for each run eight funds out of the early-stage peer benchmark group and five funds out of the later-stage peer benchmark group. After weighting the performance of each fund drawn according to a corresponding commitment size, a portfolio performance is obtained. This is repeated many times so that a distribution can be created, which is then used to benchmark the portfolio. The commitment-weighted IIRR of the portfolio is then compared against this synthetic benchmark (see Exhibit 8.12).

The results obtained from this approach should be analyzed with care. This is because, by construction (meaning the random picks), it is implicitly assumed that the fund manager knows and has access to the entire population of the peer benchmark groups. Furthermore, it is also implicitly assumed that the manager makes no decision on the allocation among the PE markets or the level of diversification, which

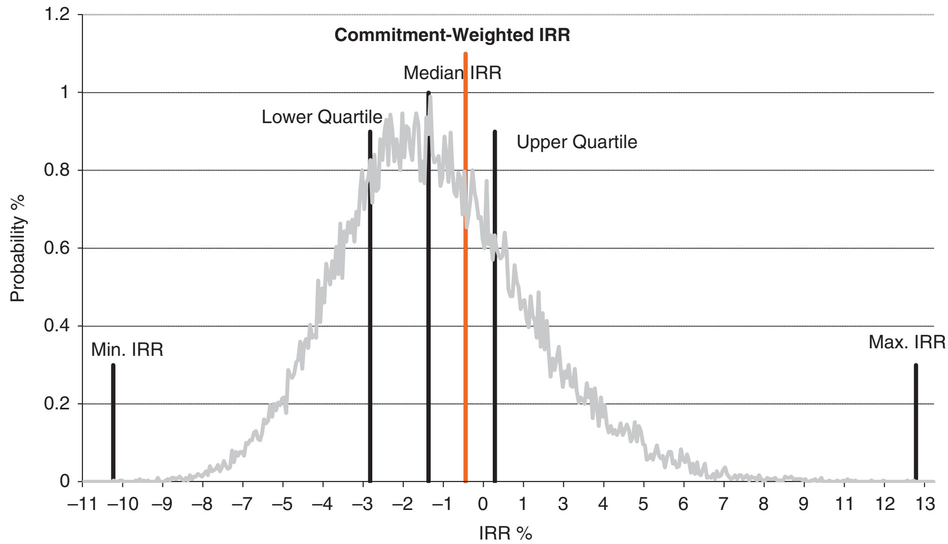


EXHIBIT 8.12 Portfolio of Funds Benchmarking: Monte Carlo Simulation

in reality is often not the case. These limitations can be resolved by running a simulation better reflecting the flexibility granted to or the constraints imposed on the manager. For example, while a portfolio can be composed of 60% buyout funds and 40% VC funds, the composition of benchmark portfolios might range between 50% and 75% of buyout funds and between 25% and 50% of VC funds, if this is what is prescribed in the investment policy imposed on the assessed manager.

8.9 CONCLUSION

The presence of a benchmark is essential for any large-scale investment program (e.g., the asset allocation program of institutional investors). Benchmarks are employed at various stages of the investment process and are used to construct strategic asset allocation models. Once a strategic allocation has been approved, the next step is manager selection. Again, benchmarks play a vital role in the evaluation process. After the allocations have been made, managers have to be monitored regularly. To understand if the manager is performing as well as expected, a benchmark along with other performance measures are employed.

To construct meaningful benchmarks, one must have measures of performance that are well defined and rooted in the fundamentals of finance. This chapter has examined various methods for measuring the performance of PE investments. These measures can also be used to compare the performance of a portfolio to two types of benchmarks. One type is constructed using peer groups, which means the benchmark is a portfolio of managers with characteristics similar to the manager that is being evaluated. Alternatively, one may use an index of publicly traded securities. In this case, various public market equivalent measures have been developed. Two such methods were discussed in this chapter.

Finally, the chapter examines the methods for measuring the aggregated performance of a portfolio of PE funds. Pooled and commitment-weighted methods are commonly used in this area.

NOTES

1. Available at www.privateequityvaluation.com. Organizations that developed the IPEV Valuation Guidelines include the Association Française des Investisseurs pour la Croissance (AFIC), the British Private Equity and Venture Capital Association (BVCA), and the European Private Equity and Venture Capital Association (EVCA).
2. See Geltner and Ling (2000): "In 1993, AIMR proposed performance measurement guidelines that recommended a time-weighted approach. After investors and fund managers expressed concerns, a special sub-committee of private equity industry investors and experts appointed by AIMR studied the applicability of time-weighted returns to the private equity industry. They recommended that fund managers and intermediaries present their private equity performance results on the cash-weighted IRR basis." CFA Institute's Global Investment Performance Standards (GIPS) and Guidance Statements are available at http://gipsstandards.org/standards/Pages/index.aspx.
3. See Chapter 16 in Mathonet and Meyer (2007).
4. Thomson Reuters and its predecessors, including Venture Economics, had been collecting and publishing buyout and venture capital return data and statistics for more than two decades. Thomson Reuters is now providing Cambridge Associates' private equity benchmarking data to subscribers to its Eikon service. See Toll (2014).
5. Phalippou (2011) finds that for emerging market spaces, the number of funds is so small that this benchmark is even less reliable than those for the mature segment.
6. A description of a Monte Carlo simulation can be found in Weidig and Mathonet (2004).

REFERENCES

Bailey, J. V., T. M. Richards, and D. E. Tierney. 1990. "Benchmark Portfolios and the Manager/Plan Sponsor Relationship." In *Current Topics in Investment Management*, edited by F. J. Fabozzi and T. Dessa Fabozzi, 71–85. New York: Harper & Row.

Cumming, D. J., L. H. Hass, and D. Schweizer. 2013. "Private Equity Benchmarks and Portfolio Optimization." April 15. http://ssrn.com/abstract=2251001.

Geltner, D., and D. Ling. 2000. "Benchmark and Index Needs in the US Private Real Estate Investment Industry: Trying to Close the Gap." RERI Study for the Pension Real Estate Association.

Mathonet, P.-Y., and T. Meyer. 2007. *J-Curve Exposure: Managing a Portfolio of Venture Capital and Private Equity Funds*. Chichester, UK: John Wiley & Sons.

Mathonet, P.-Y., and G. Monjanel. 2006. "Valuation Guidelines for Private Equity and Venture Capital Funds: A Survey." *Journal of Alternative Investments* 9, no. 2 (Fall): 59–70.

Nikulina, L., M. Calnan, and G. Disdale. 2012. *Private Equity Benchmarking: Where Should I Start?* https://www.towerswatson.com/en-US/Insights/IC-Types/Survey-Research-Results /2012/03/Private-Equity-Benchmarking-Where-Should-I-Start.

Phalippou, L. 2009. "Risk and Return of Private Equity: An Overview of Data, Methods and Results." Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1352351.

———. 2011. "An Evaluation of the Potential for GPF G to Achieve Above Average Returns from Investments in Private Equity and Recommendations Regarding Benchmarking." Report for the Ministry of Finance and the Norwegian Parliament, February.

Swamy, G. M., I. Zeltser, H. Kazemi, and E. Szado. 2011. "Setting the Benchmark: Spotlight on Private Equity." *Alternative Investment Analyst Review*, CAIA.

Swensen, D. F. 2000. *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*. New York: Simon & Schuster.

Toll, D. M. 2014. "Thomson Reuters Partners with Cambridge Associates on Benchmark Data." March. https://www.pehub.com/2014/03/thomson-reuters-partners-with-cambridge-associates-on-benchmark-data/.

Weidig, T., and P.-Y. Mathonet. 2004. "The Risk Profiles of Private Equity." Working paper. EVCA, Brussels, January.

CHAPTER 9

Fund Manager Selection and Monitoring

While a wide divergence between top- and bottom-quartile performers may provide an opportunity to perform extremely well by selecting a top-performing manager, the large dispersion in manager returns may also expose a portfolio to a high degree of underperformance risk.

For individual transactions, comparison against a benchmark is not only a key part of the due diligence process through track record analysis, but also a key performance indicator during the monitoring phase. The **monitoring phase** of private equity (PE) fund commitments takes place at the fund level and focuses on financial performance and the manager's structural and behavioral developments. The manager's structural and behavioral developments include adherence to the strategy and governance structures, compliance with the terms of the partnership agreement, and contribution of value to the underlying portfolio companies.

9.1 PERFORMANCE PERSISTENCE

There has been a wide consensus among investors that in private equity, there is a link between prior fund returns and those in the future. This belief implies that limited partners (LPs) should allocate their money to general partners (GPs) that have demonstrated top-quartile performance in the past. While a GP's competencies remain critical in the success or failure of a fund, these competencies played a more significant role in the early days of the private equity industry. The industry has now matured significantly, with knowledge and experience spread widely, so skills alone are now less likely to have a huge impact.

Although private equity GP performance persistence is widely accepted as a fact, recent research has put this into question. These studies have shown that while persistence exists for certain strategies, the degree of persistence has weakened considerably since 2000.[1] For example, the performance persistence of follow-on funds relative to the performance persistence of the previous funds has weakened since 2000.

Going forward, LPs cannot rely on simple screening criteria any longer and need to reflect on how they can create value under these circumstances. This also implies greater focus on aspects of manager selection and monitoring that form distinct parts of the investment process, that can be efficiently structured, and that are among the keys to sustainable outperformance in private equity.

EXHIBIT 9.1 Transition Matrix of Private Equity Funds, 1980–2003

		Next Fund		
		Lower 3rd	**Middle 3rd**	**Upper 3rd**
Current Fund	**Lower 3rd**	43%	29%	28%
	Middle 3rd	30%	41%	29%
	Upper 3rd	26%	32%	42%

Source: Phalippou and Gottschalg (2009).

9.1.1 The Seductive Case for Performance Persistence

GPs' performance persistence in private equity has long attracted the interest of practitioners as well as academics.[2] What evidence there is seems to support the performance persistence hypothesis.

Available empirical evidence strongly supports the presence of persistence for all private equity strategies until the late 1990s. For venture capital (VC), Conner (2005) found that follow-on funds raised by fund managers whose prior fund's performance was in the top quartile had a 44% chance of achieving a top-quartile internal rate of return (IRR). If the distribution of returns across GPs was perfectly random and previous returns were irrelevant, this figure would be 25%. Hendershott (2007) suggests that when a manager has had three consecutive top-quartile PE funds, potential investors can conclude with 80% or greater confidence that the manager is exceptional. The various studies tend to agree that two successive funds have positively correlated returns, for both buyouts and venture capital.

Exhibit 9.1 is adapted from Phalippou and Gottschalg (2009) and presents transition probabilities concerning the performance of private equity funds (both buyout and VC) using data covering 1980–2003.

According to these results, there is a 43% chance that the follow-on fund of a fund that was ranked in the lower third will be in the lower third as well. On the other hand, there is a 42% chance that follow-on funds of top-tier funds will be in the top tier as well.

When considering private equity's modus operandi and the observation that GPs have differential and proprietary skills, the performance persistence hypothesis appears to be highly plausible:

- Distinct organizational capabilities could credibly explain the top-performing GPs' superior value creation.[3]
- Such GPs are supposed to be building on their extensive industry networks, which give them an advantaged deal flow and allow them to identify the most promising deals.
- Associated with such a network is domain expertise (that is, specialized knowledge of the industry sector in which the GP is operating).
- Finally, and largely independent from industries, top-performing GPs are said to have the ability to create value through operational improvement in portfolio companies.

Consequently, gatekeepers market their benchmarking services for checking and comparing GPs' performance as well as their ability to sell access to these "top team" fund managers. **Gatekeepers** are professional advisers operating in the PE market on behalf of their clients: in particular, consultants and account managers, funds of funds, and placement agents.[4] The fact that the performance persistence argument is neat, simple, and intuitively sensible means any result that conflicts with this conventional wisdom tends to be met with skepticism.

On a theoretical level, significant performance persistence poses a problem as well, especially given the anecdotal evidence that some top-performing funds do not accept all the money that investors are willing to provide. Why don't these fund managers increase their income by increasing fund size or fees? Kaplan and Schoar (2005) argue that, given the diseconomies of scale displayed by private equity funds, top-performing funds limit their size so that they can deliver higher returns to their investors. Higher performance will enable these fund managers to raise additional funds going forward. Marquez, Nanda, and Yavuz (2010) argue that, unlike mutual funds that invest in public securities, investments by private equity funds are subject to a two-sided matching problem: on one side, private equity funds want to invest in high-quality entrepreneurial firms; on the other side, entrepreneurs seek to match with talented fund managers that are likely to add value. Further, firms are more likely to accept offers from VC firms that are more reputable and have a higher ability to add value. Finally, private equity funds want to attract a certain type of investor, such as those who will have the funds to respond to capital calls and to invest in follow-on funds.

9.1.2 Six Challenges to the Performance Persistence Hypothesis

However, there has been increasing skepticism in recent years regarding the persistence hypothesis. While pre-2000 data support the hypothesis for both buyout and VC firms, the evidence is much weaker for buyout funds post-2000, except at the lower end of the performance distribution. Performance of the current fund was found to be statistically indistinguishable regardless of the quartile positions of the GPs' previous funds.[5] Persistence has declined in the 2000s relative to the 1990s.[6] It is found to be short-lived, with best- and worst-performing funds performing similarly in their second and third follow-on funds, and performance converging across funds over time.[7] Indeed, on close examination, performance persistence in private equity is an inadequate credo. There are six primary challenges to the performance persistence hypothesis:

1. Ambiguities regarding definition of "top performance": Conscious of the widespread belief in performance persistence, LPs' preoccupation with track record, and LPs' tendency to meet only with "top-quartile" managers, GPs frame the measurement of their past performance in whatever way will demonstrate that they are indeed "top." Although the quartiles' mathematical definition is clear, what the numbers refer to is vague. Do they refer
   - to the funds' internal rates of return or to their multiples?
   - to the funds' interim or final returns?

- to the funds' gross returns or to LP returns net of fees and carried interest?
- to which peer group? Do they refer to private equity as a whole or do they differentiate between, say, buyouts and venture capital?

2. Comparing apples and oranges: If we base our definition of success on the quartile position in the benchmark, comparison should be made against funds that are subject to the same market conditions in order to assure that apples are being compared to apples; that would be the vintage-year cohort of the peer group. However, from vintage year to vintage year, the composition of the peer group will change; rarely can two funds managed by the same management team be measured against the same peers. GPs raise funds in irregular intervals; therefore, the firms that raised the funds that made up the previous vintage year's peer group may not be out looking for investors at the same time for their next round of fundraising. Drawing the conclusion that success as measured by quartiles is primarily a result of the GPs' skills assumes that all funds had equal starting positions and faced an identical group of competitors.
3. Luck or skill? Performance may depend on talented individuals or on the processes and culture of the GP as an organization. While it seems an accomplishment to be in the top quartile, in a 30-year career and even assuming that the fund management team's composition does not change, a GP will have only a handful of opportunities to manage funds over the GP's full lifetime. Assuming that a manager can start a new fund every six years, the manager may have only four funds that have been completely liquidated and that therefore offer unambiguous performance figures. This is hardly enough data to estimate the impact of the manager's skills with any confidence.[8]
4. Growth in fund size: First-time managers tend to improve their performance in successive funds; however, this is insufficient as evidence for learning and increasing skills.[9] Research findings suggest a concave relation between fund size and performance: up to the point when their performance tends to decline again, larger funds also have higher returns.[10] As first-time funds are usually smaller than the average, the performance improvement is more likely to be linked to the growth in fund size.[11]
5. Secular market trends: In situations in which all participants make money at the same time (for example, in the VC market at the end of the 1990s), performance is likely to be driven by a secular trend rather than by individual talent. There is evidence that relative fund performance is, among other factors, a function of industry selection.[12] This is particularly true in situations in which top-quartile funds choose industries at a stage of a cycle that has a major positive impact on their subsequent returns.
6. Statistics relate to different niches in the PE market: For PE funds, the difference between a top-quartile return and a median return can be two or three times as much as in the case of mutual funds, for which persistence has generally been found to be very weak.[13] However, mutual funds are not the right reference for comparison, as they are essentially operating within the same level playing field: efficient public markets. As an analogy, we can look at public markets as everybody swimming in the same river; thus, differences in the swimmers' speeds are likely to be explained by their strength. PE fund performance figures, in contrast, are relating to different segments in the real economy with

EXHIBIT 9.2 Transition Matrix of Buyout Fund Performance, 1984–2000

Buyout Funds		Current Funds (pre-2001)			
		1	2	3	4
Previous Fund Quartile	1	37.5%	25.0%	18.8%	18.8%
	2	30.4%	21.7%	30.4%	17.4%
	3	21.4%	25.0%	32.1%	21.4%
	4	17.4%	26.1%	30.4%	26.1%

Source: Harris, Jenkinson, and Stucke (2014).

different dynamics. In private equity, we are comparing the speed of swimmers in different rivers with different currents, where the conditions of each river are more likely to be the most important factor in each swimmer's speed. It is therefore unlikely that it is mainly skills that lead to the significant differences in GP performance.

Actually, it is difficult to see why GPs should be exempt from the rule that all empires eventually decline and fall and why GPs (in contrast to, say, hedge fund managers) should be able to sustain their excellence instead of being exposed to the dynamism of markets, where all competitive advantages are temporary.[14]

9.1.3 Performance Persistence Hypothesis and Transition Matrices

Exhibit 9.2 and Exhibit 9.3, which are adapted from Harris, Jenkinson, and Stucke (2014), display the transition matrix of buyout funds for pre-2001 and post-2000 periods. The data cover 1984 through 2011.

Exhibit 9.2 clearly shows that before 2001, there was significant performance persistence among the top-tier buyout funds. For example, a top-quartile fund had a 37.5% chance to be in the top quartile of the performance of its follow-on fund.

EXHIBIT 9.3 Transition Matrix of Buyout Fund Performance, 2001–2011

Buyout Funds		Current Funds (post-2000)			
		1	2	3	4
Previous Fund Quartile	1	22.0%	28.8%	30.5%	18.6%
	2	24.5%	22.6%	32.1%	20.8%
	3	15.4%	28.2%	38.5%	17.9%
	4	21.4%	14.3%	32.1%	32.1%

Source: Harris, Jenkinson, and Stucke (2014).

EXHIBIT 9.4 Transition Matrix of Venture Capital Fund Performance, 1984–2011

VC Funds		Current Funds (whole sample)			
		1	2	3	4
Previous Fund Quartile	1	48.5%	16.7%	24.2%	10.6%
	2	28.9%	34.2%	20.2%	16.7%
	3	22.0%	29.4%	29.4%	19.3%
	4	14.8%	17.3%	29.6%	38.3%

Source: Harris, Jenkinson, and Stucke (2014).

A bottom-quartile fund, in contrast, had an 18.8% chance of finding its follow-on fund to be a top-quartile fund.

Exhibit 9.3 shows that the degree of performance persistence has declined, especially for top-quartile managers. However, we continue to see strong persistence for poor-performing funds. A bottom-quartile buyout fund had a 32.1% chance of seeing its follow-on fund ranking in the bottom quartile, whereas there was an 18.6% chance of the same thing happening to a previously top-quartile manager.

Exhibit 9.4 displays the transition matrix for VC fund performance covering the entire sample. We are not reporting the results for two different sample periods, because, unlike the results for buyout funds, the results for VC funds have remained rather consistent. For example, a top-quartile VC fund had a 48.5% chance of seeing its follow-on fund appearing as a top-quartile fund for the whole sample. The same figures for the split samples were 48.7% and 48.1% for pre-2001 and post-2000 periods, respectively.

It is important to note that these results are not universally supported, as other studies report different figures for the transition matrix. The differences are primarily driven by the variety of databases that are employed and the adjustments that researchers make in order to improve the integrity of the data. For example, some researchers assume that the book values of mature funds that have been partially liquidated and that have remained inactive for several years is zero, while others accept the values reported by the fund manager.

For a more complete picture, Exhibit 9.5 reproduces a chart provided by Preqin, a well-known source of data in the alternative industry. The chart covers the overall private equity fund universe for 1982–2012.

We can see that over the period covered by the study, 36% of top-quartile funds produced follow-on funds that ranked in the top quartile in terms of performance. On the other hand, only 13% of top-quartile funds produced follow-on funds that ranked in the bottom quartile.

These and other results point to the follow broad conclusions:

- Performance persistence is present for poor-performing buyout and VC funds. The persistence has weakened in recent years but is still economically significant.

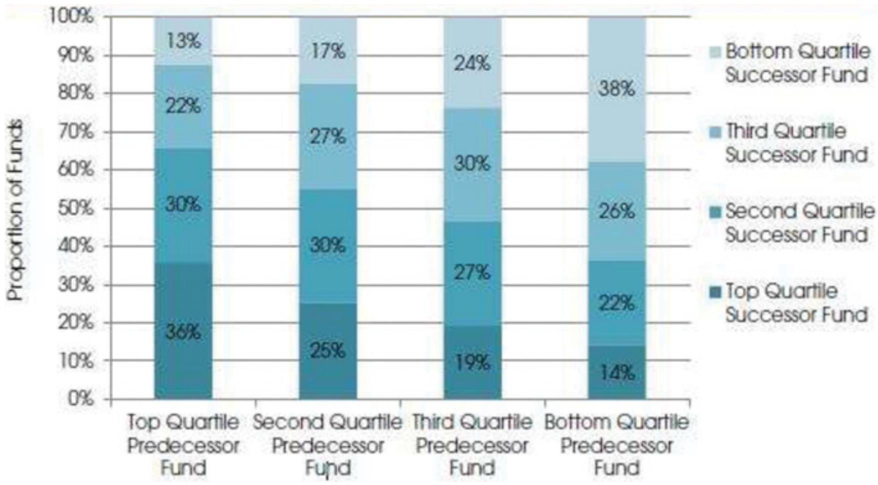


EXHIBIT 9.5 Overall Private Equity: Relationship between Predecessor and Successor Fund Quartile Performance
Source: Preqin Private Equity Report (2016).

- Performance persistence has almost disappeared for top-performing buyout funds in recent years.
- Performance persistence has weakened slightly for top-performing VC funds in recent years but remains economically highly significant.

9.1.4 Performance Persistence Implementation Issues

Even if investors believe in the performance persistence hypothesis, in practice they will find it difficult to make meaningful use of it: by the time there is sufficient evidence that a firm is really successful, the effect is likely to be gone. The main problem is that all analyses on performance persistence in private equity are based on data for funds close to the end of their lifetime. Only funds aged between at least six and eight years are sufficiently mature for drawing conclusions regarding performance, but GPs need to raise subsequent funds in years 2 or 3.[15] In other words, top-quartile performance is not reliably visible before investors have to make their next commitment. All these factors make it difficult for investors to use top-quartile performance as an effective screening criterion.[16]

9.2 MANAGER SELECTION AND DEAL SOURCING

Manager selection is not mechanical but requires industry experience and resources to conduct both research and due diligence. Unfortunately, this is easier said than done, and the advice to focus on top funds is probably as helpful as the observation that to become rich, one needs to acquire a lot of money. Further, it is more difficult

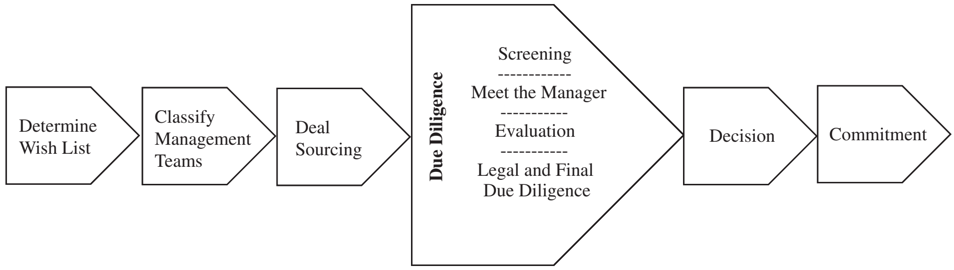


EXHIBIT 9.6 PE Fund Manager Selection Process

to identify superior managers than it is to weed out obviously inferior managers. The key is to be highly selective and to strike a proper balance between seeking exposure to top funds and diversifying. Thorough, consistent, and detailed analysis and discipline in the due diligence process are critical.

To make matters worse, few investors, advisers, or consultants have experience and familiarity with the unique aspects of private equity. Because the industry and its practices are continuously evolving, categorizations are fuzzy and there are no clear dos and don'ts. That makes the identification and evaluation of fund managers both more important and more challenging in the PE universe. Consequently, one needs a different selection process for fund managers in private equity (see Exhibit 9.6) than for fund managers of publicly quoted assets.

9.2.1 Determination of the Wish List of Fund Characteristics

The development of an investment strategy is important to efficiently manage the process, and it forms the starting point of the fund manager selection process. Based on the investment strategy of the investor and the resulting portfolio design, a wish list of fund characteristics needs to be established. This wish list defines the types of proposals that are consistent with the investment strategy of the investor.

Next, an active deal sourcer identifies wish-list funds to be specifically targeted for investment. Investors make a market mapping, in which all management teams are ranked by their perceived attractiveness (see Exhibit 9.7).

9.2.2 Classifying Management Teams

Despite the reservations regarding the performance persistence hypothesis, track record remains an important factor to take into consideration when assessing a fund management team's competence. Attractive teams are normally those that have been able to generate top performance during several market cycles, as they are the most likely to continue to do so in the future. One way to classify manager teams is to rank them according to two dimensions: (1) the quality of their **track record** (i.e., their financial performance in previous ventures) from bottom to top performer, and (2) the duration of their **joint experience** (i.e., their previous involvement as

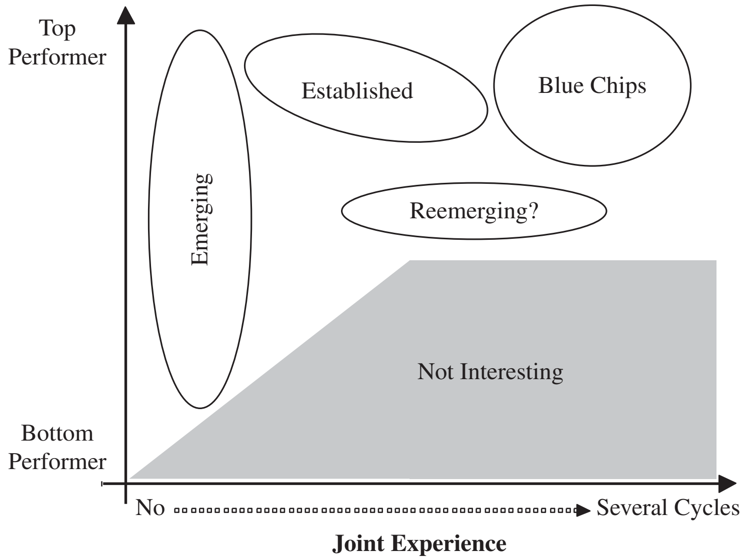


EXHIBIT 9.7 Market Mapping of Fund Performance through Several Business Cycles

a management team) from none to several market cycles. From this, the following classification is suggested:

- A **blue-chip team** is a team that has been able to generate a top-quartile performance for all of its funds through at least two business cycles (i.e., a sequence of more than three funds).
- An **established team** is a team that has been able to generate a top-quartile performance for most of its funds (more than three funds) through at least two business cycles.
- An **emerging team** is a team with limited joint history but with all the characteristics to become an established team.
- A **reemerging team** is a previously blue-chip or established team that has been through a restructuring following recent poor performance or some significant operational issues and has regained the potential to reemerge as an established or a blue-chip team.

Teams not included in the preceding categories are not usually considered by most investors. Although in specific situations (for example, in emerging markets) the overall market opportunity may be very interesting, it will not be possible to find teams with track records comparable to those of teams in more established PE markets.

9.2.3 Deal Sourcing

Investors in PE funds need to use their network of industry contacts to identify and establish communication with high-quality fund managers. It is critical to get as many opinions and leads as possible; this can be achieved by having discussions with other investors or entrepreneurs, employing advisers and consultants, and researching the press.

Reactive deal sourcing, in which investors evaluate a portion of the large number of general partners who contact them to ask for an investment, is not an efficient way of approaching selection. With this approach, literally hundreds of private-placement memoranda need to be checked as to whether they comply with the set wish-list characteristics or investment criteria. First-time teams often have to approach as many investors as possible, but top teams typically get referred to limited partners by word of mouth and therefore need to be actively sourced. Developing long-term relationships and exploring opportunities that fit the set criteria are critical. For this purpose, teams have to be identified and approached even before they start their fundraising. This process requires establishing a calendar of when teams are expected to go to the market to raise capital for their next fund offering.

Under the performance-persistence assumption that past performance and future performance of a team are correlated, funds raised by teams that have performed well in the past tend to be oversubscribed. If returns have been high, LPs from prior funds are highly motivated to commit to follow-on funds, and GPs will reward their loyalty with virtually guaranteed access to future funds.

In fact, the PE industry is a close community in which past relationships provide access to funds in high demand, while newcomers are often turned away. Frequently, when a team with a strong track record raises a new fund, investors in previous funds quickly commit, often leading to oversubscription. GPs are interested in maintaining their relationships with the existing LP base. Searching for new LPs is an expensive exercise, and it creates uncertainty regarding the timing of closing and future relationships.[17] Consequently, fund managers tend to avoid this exercise whenever possible. In the extreme, it will not even be known to outside parties that the team may be raising a new fund. Having access to a network of contacts is required to identify top teams and to know about the timing of their fundraising activities. While top teams give priority allocations to their previous investors, they may also allocate a share of the new fund to investors who could add value, such as deal flow, exit opportunities, and industry expertise. Nonetheless, access is far less a problem for LPs who are financially strong and have demonstrated that they are long-term players in the market. For newcomers, however, this is a significant barrier to entry. This means that an investor building a new allocation to private equity may not be able to invest with the top teams, even when the investor has knowledge of their fundraising schedules.

9.3 DECISION-MAKING AND COMMITMENT

Due diligence can be seen primarily as information gathering and evaluation and not as a decision-making tool. In practice, the distinction is seldom made; due diligence is used to weed out inferior funds and to accept the remaining proposals. The results of the due diligence process should be used only as input for a decision-making process that takes into consideration not only the quality of the investment proposal but also the program's portfolio composition.

Finally, this is not a one-sided decision. Teams may have their own due diligence criteria for selecting potential investors. They should examine whether the investor's commitment is long-term, whether the investor understands the business, whether the investor has a reputation for being difficult, and whether the investor has been a **defaulting investor** (one who has previously reneged on capital commitments) or is at risk of defaulting due to financial distress. In a case in which public institutions seek to become LPs, their investment restrictions (particularly those related to industry sectors and geography) and their transparency requirements need to be acceptable to the fund managers. In order to protect the GP's private information, some PE fund managers may not allow commitments from investors who are required to publicly disclose information regarding their investment programs, such as U.S. pension plans subject to the Freedom of Information Act.

9.4 PRINCIPLES OF FUND MONITORING

Investors in PE funds may assume that little can be done, besides divesting on the secondary market, to prevent problems once the due diligence process has been completed and the commitment to the fund has been made; however, ongoing monitoring throughout the life of a PE investment is a necessary control mechanism. In such a long-term business, initial due diligence findings quickly become obsolete, while changes to the economic environment can fundamentally alter the balance between investor and fund manager interests. The information asymmetry and moral hazard–related problems associated with such changes can be lessened through monitoring.

Monitoring involves the routine and systematic collection of information. Robbie, Wright, and Chiplin noted in 1997 that private equity LPs typically engaged in few monitoring actions. The authors expected this trend to continue, as a more proactive approach often raises questions about cost-effectiveness.

9.4.1 Monitoring as Part of a Control System

Monitoring involves more than simply the issuance of warnings. Instead, it should be seen as part of a larger control system within the investment process (see Exhibit 9.8). Its role is to observe, verify, and control in an attempt to make the portfolio perform in a desired way.

The monitoring process involves identifying problems and developing a plan to address them. Because of the illiquidity of a PE fund, the investor's ability to react to

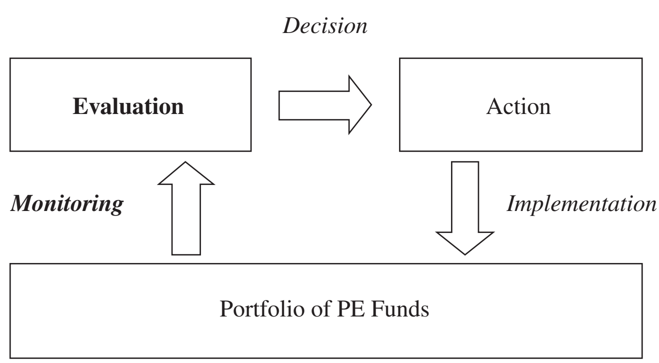


EXHIBIT 9.8 Control System

infrequently identified problems is somewhat limited. In many situations, a solution will require finding a consensus with the fund manager and the other LPs, or building alliances with the LPs to exercise pressure and act jointly. This explains why due diligence should focus not only on the fund manager but also on the other LPs.

9.4.2 Monitoring Trade-Offs

In many ways, the limited partnership agreement's terms and conditions reflect a trade-off in various dimensions. It is neither possible nor meaningful to anticipate the expected behavior of a GP over a period of 10 or more years. Market conditions fluctuate, fund management teams evolve, co-investors change, investments do not materialize as planned, and new opportunities arise. In hindsight, LPs may conclude that certain provisions of the agreement were overly restrictive and worked against their interests, while other issues may not have been addressed in sufficient detail. The monitoring exercised by investors is intended to bridge this gap.

Investing in and monitoring PE investments involves more effort and higher costs relative to an otherwise equivalent publicly traded investment. Such costs need to be weighed against potential benefits. Turning the monitoring findings into management actions, as well as choosing the appropriate time and degree of intervention, poses a dilemma. This is because intensive monitoring and its associated interventions can effectively dilute the fund manager's responsibility and lead to LPs being reclassified as GPs and, in extreme cases, negate their limited liability status. Private equity, in particular, requires the long-term view, and LPs should avoid overreacting to bad news, especially during the early years of a fund's lifetime. However, the PE investor's reaction, when necessary, typically comes too late to prevent further deterioration of the investment.

9.5 MONITORING OBJECTIVES

While monitoring is an important instrument for ensuring compliance with the limited partnership agreement's terms and conditions and for gathering information, it is not necessarily linked to the performance of a fund. Monitoring should not be confused with the management of portfolio companies, an activity in which, in accordance with industry practices, LPs have no involvement. Rather, LPs are responsible for managing the portfolio of funds and monitoring fund managers.

9.5.1 Portfolio Management

Investments into funds are made in the context of a general portfolio strategy. Because of the blind-pool nature of PE fund investing, it is crucial for LPs to set the risk profile of their investment at the time of commitment. Moreover, given private equity's lack of liquidity, LPs cannot easily adjust portfolio holdings or rebalance them if the GP undertakes actions that are inconsistent with governing documentation.

Consequently, monitoring must encompass not only individual funds but also the limited partner's overall portfolio composition. Indeed, while a detailed asset allocation process can promote effective diversification and eliminate the problems associated with haphazard fund selection, in the case of private equity it is often problematic to obtain equal weightings across funds and, at the same time, be represented in all key market segments. In addition, significant changes in investment valuations due to market fluctuations and distribution activity may cause the allocation to private equity to rise above or fall below desired limits. To ensure that the allocation remains within the established range, LPs must consistently monitor and adjust the portfolio structure.

LPs monitor fund managers to control risk. In conventional asset classes, reducing risk means moving money into safer investments. But unlike most investors, LPs cannot easily withdraw their commitments. Through monitoring activities, however, an LP may be able to identify severe shortcomings in time to reduce downside risks, either through restructuring or by selling the position on the secondary market.

9.5.2 Monitoring the Risk of Style Drift

Monitoring is also important for ensuring style discipline. Because LPs are investing in a blind pool, the investment is based mainly on the fund manager's declared investment strategy. That said, there are risks associated with adhering too closely to a declared investment strategy, especially when market conditions change significantly, creating new opportunities. Therefore, funds will not necessarily adhere to it, nor should they in certain instances.

For example, in a difficult market environment, it makes sense to deviate from the declared investment strategy and to look for investments in more promising areas. While style tracking in private equity is not as applicable as it is in the context of the hedge fund industry, LPs nevertheless need to ensure that fund managers stay within the confines of their core expertise and style. LPs should be concerned with any change to a PE fund strategy. Indeed, style drift may have serious consequences for the risk-return profile of the fund and can create unexpected exposure for LPs. This reinforces the need to constantly monitor and adjust the portfolio structure.

Style drift describes the tendency of investment managers to deviate over time from their initially stated and agreed-on investment style. In private equity, fund managers explain their investment practices and their strategies during the initial due diligence process. LPs expect fund managers to be reasonably consistent in following these practices. However, adherence to a stated investment style may not always hold true in the world of PE funds, where secrecy and flexibility are critical to success. In researching this issue, Cumming, Fleming, and Schwienbacher (2004) find that style drift in private equity is more common than was previously perceived.

Changes in style are often observed in geographical focus, between buyout and venture capital focus, or by targeted industries. The skill sets required for fund managers and the investment objectives of buyouts and VC funds differ in important ways. While related, the two fields are sufficiently distinct to make the transfer of skills difficult. Nevertheless, there is anecdotal evidence that VC funds that raised excessive amounts of money were not able to resist the temptation to put the money to work in buyouts, rather than return the unutilized commitments to investors. When deal flow dries up, fund managers often consider other markets. For example, U.S. venture capital funds may look for investments in Europe, while European venture capital funds try to gain access to Silicon Valley. LPs view this geographic drift with skepticism because, particularly in the case of venture capital, hands-on involvement of the fund manager is essential. Moreover, with the change in geography, investors may become exposed to foreign exchange risks they had not accounted for previously. In order to alleviate the risk of style drift, the up-front design of the limited partnership agreement is important, as its covenants guide the behavior of the fund manager.

9.5.3 Creating Value through Monitoring

For fund investments, the management of the upside is primarily delegated to fund managers, assuming that appropriate incentives have been provided. This underscores the importance of selecting the right teams. While the GP is able to create value at the individual PE fund level, LPs can create significant value through monitoring activities at the portfolio of funds level, as illustrated in the following five situations:

1. The study undertaken by Lerner, Schoar, and Wongsunwai (2007) suggests that investors in private equity owe their success to superior reinvestment skills. The authors specifically refer to the example of endowment funds. These funds were found to be less likely to reinvest in a partnership, but if they did invest in the follow-on fund, its subsequent performance was significantly better than those of funds they let pass. This finding underscores the importance of monitoring for improved decision-making.
2. Intensive contact with the fund managers is important when deciding whether to invest in a follow-on fund (i.e., re-ups). It improves the due diligence process and can lead to a quicker finalization of contracts after incorporating improvements based on previous experience with the fund manager. Moreover, a strong relationship can extend to junior team members ready to spin out and set up their own fund.
3. Networking and liaising with other LPs is an important instrument for gathering intelligence on the overall market and gaining knowledge of other funds, and may help an investor gain access to deals that might otherwise not appear on the institution's radar screen.[18] It can also improve access to secondary opportunities in advance of the less favorable auction process.
4. In the context of a co-investment strategy, monitoring is important for screening interesting investment opportunities that may arise.
5. Access to information may enable an LP to optimize the management of commitments through more precise cash flow forecasting.

Finally, lessons learned from monitoring can be applied in the future to improve the due diligence process and the selection of future investments.

9.6 INFORMATION GATHERING AND MONITORING

The private equity sector is called "private" for good reason, and transparency has its limitations. Transparency is the degree to which investment holdings and strategies are disclosed to investors. The typical monitoring process follows a dual approach (see Exhibit 9.9), separating formal from informal reporting. There is a tendency for larger investors to differentiate between obtaining specific qualitative data by direct interaction with investment managers and obtaining quantitative or standardized

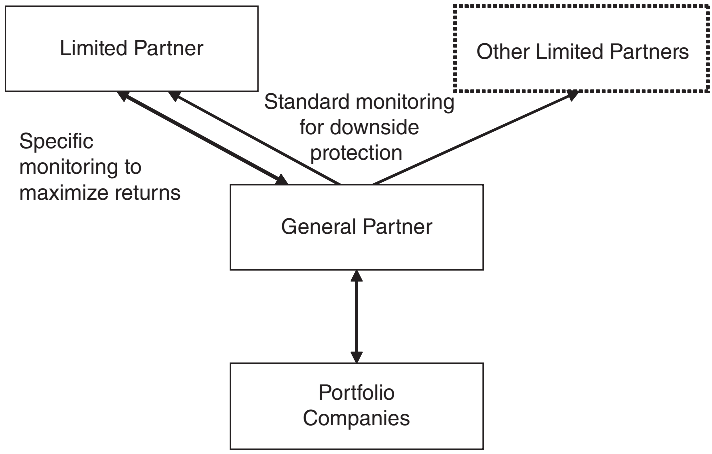


EXHIBIT 9.9 Monitoring Process

data provided by the back office. As reporting quality and detail vary considerably among different funds, monitoring needs to focus on filling the gaps present in the reporting. To avoid the risk of reporting overloads, an appropriate balance must be struck between the provision of specific information and the provision of standardized information to the LPs.

9.6.1 Transparency

Although reporting to investors can often be more transparent for PE investments than it is for public ones, this transparency is normally kept to a minimum for noninvestors. Thus, the development of valuation and reporting standards in private equity has been driven more by industry players than by regulatory bodies, although this has to some degree changed with a tightening of regulatory regimes for alternative investments worldwide.

Historically, the appearance of private equity as unregulated was the by-product of a careful balancing of specific exemptions from certain aspects of regulation and the sustained efforts of PE and VC associations to maintain high standards of investment conduct by private equity fund managers. In the United States, PE fund managers and hedge fund managers have historically relied on the same exemption from registration under the Investment Advisers Act of 1940.

The information required for the management of a portfolio of PE funds is, in principle, the same as that needed for the investment decision (i.e., the information gathered through the due diligence process). This information is supplied by the fund manager but is also collected from outside sources.[19] While a significant amount of the information should be provided in a low-cost, standardized manner, proprietary information can lead to a competitive advantage.

9.6.2 Standard Monitoring Information

Monitoring is built around the information supplied in fund management reports. Therefore, reliable, appropriate, and timely management of information is crucial. PE associations' reporting and valuation guidelines, such as those issued by the Institutional Limited Partners Association (ILPA) and the International Private Equity and Venture Capital (IPEV) Valuation Board, define the standard level of information to be provided. Generally, investors in PE funds complain about inadequate disclosure, inconsistent reporting, and the often-significant time lag in receiving this information. Major problems with a portfolio company eventually appear in the reporting. Often the damage has been done by then, and little corrective action remains possible. This raises the question of whether large institutional investors will become more active or proactive in addressing the following issues:

- Incomplete information creates uncertainty and consequently increases risks. Should institutions, possibly together with PE associations, persuade industry players to provide more detail and quality in reporting?
- Transparency in reporting is tied to the question of valuation and risk. Can investors get a truer and fairer picture of their investments through more detailed and better-quality investment reporting?
- There is overall agreement among industry players that formal accounting-related information should be standardized. Industry reporting guidelines are seen as sufficient in general, but the information reported greatly depends on the GP's willingness to disclose this information in the reporting framework.
- Increased disclosure appears to be a desirable objective, especially to those accustomed to public markets. However, it needs to be kept in mind that there are practical obstacles to disclosure. The answers to these questions need to be pondered against the background of the PE market's dynamics and from a cost-benefit viewpoint.
- In private equity, information collection and analysis can be prohibitively expensive, and there are market forces that work against transparency.[20] Any attempt to go against these dynamics would be expensive and time-consuming and might ultimately be unsuccessful.
- The huge information asymmetry in private equity explains and justifies the need for intermediation. An increased level of detail will not allow a nonexpert to verify a fund manager's reporting and appraisal. Moreover, due to its imprecise character, any qualitative information is, by definition, of low quality. If one does not want to rely on the intermediary's judgment, the only alternative is direct investing, which requires an entirely different approach and skill set.
- Because PE fund investments are buy-and-hold and illiquid, a fund's LPs are unable to react quickly to new information. Moreover, for such a long-term-oriented investment, short-term developments do not, in most cases, materially affect the fund's valuation.

Based on these points, one could conclude that although a higher degree of transparency would be desirable, the public market cannot be seen as the benchmark. It is mainly the LPs' monitoring that can overcome the lack of transparency and the reporting time lags of this asset class, but only to the extent that the general manager is willing to cooperate toward enhanced disclosure.

9.6.3 Specific Information

GPs are extremely reluctant to disclose all information to investors. Their dilemma is obvious: on the one hand, there is an obligation to disclose information so that investors are able to understand the portfolio's progress; on the other hand, further information, especially at a level of detail that allows an independent risk assessment, may potentially increase the chance that LPs will start investing directly and not commit to follow-on funds. There is also the investment rationale for maintaining a high degree of confidentiality. Indeed, a fund with a niche strategy that consistently yields above-average returns will attract competition. GPs fear that too much information given to the outside helps competitors imitate their strategy, access their deal flow, or jeopardize their negotiating position. Cullen (2004) cites an example in which an LP was given financial information regarding a deal that was later shared with a competitor and ultimately caused the fund to lose the deal.

Moreover, if disclosed to a wider audience, information can be highly damaging. It may even have an adverse impact on the trading ability of a portfolio company, as it could result in reduced credit lines or lead the company's potential clients to choose to partner with the competition. Conversely, news of success may breed competitors. In the extreme, the fund manager might even be sued for disclosing harmful information.

Interestingly, however, it is not only the fund managers but also the LPs who may prefer to limit the degree of transparency. Making top teams and their investment focus public knowledge may attract competitors. LPs need to protect their privileged access to follow-on funds or to new teams that set up their own vehicles outside the old fund. PE funds are not scalable; therefore, LPs may be concerned about being locked out of follow-on funds because, as suggested by Lerner and Schoar (2004), GPs have a preference for deep-pocketed investors. It is better to keep the information quiet than to attract a feeding frenzy of competing investors.

9.7 ACTIONS RESULTING FROM MONITORING

The results of monitoring and its evaluation can lead to decisions on a series of possible actions. These can range from simply changing the monitoring intensity to intervening at the individual fund or perhaps even at the portfolio company level. The monitoring intensity should be a function of the total exposure of a PE fund, its final expected performance, and its operational status (see Exhibit 9.10).

It does not make sense to spend too much time on monitoring funds that are already quite advanced in their life cycle or beyond recovery. It is also not money well spent to focus scarce monitoring resources on teams that are highly professional or on funds over which other experienced LPs already exercise significant oversight. Finally, the cost of control must be relative to the size of the asset. Active involvement and other actions mainly relate to individual funds and can take seven forms: three forms of active involvement in the fund's governance process, two actions outside the fund's governance process, and two actions outside fund investing.

9.7.1 Three Forms of Active Involvement in the Fund's Governance Process

The three forms of active involvement in the fund's governance process are:

1. During the lifetime of a fund, agreements are not carved in stone. If it becomes clear that the original investment strategy cannot be successfully implemented

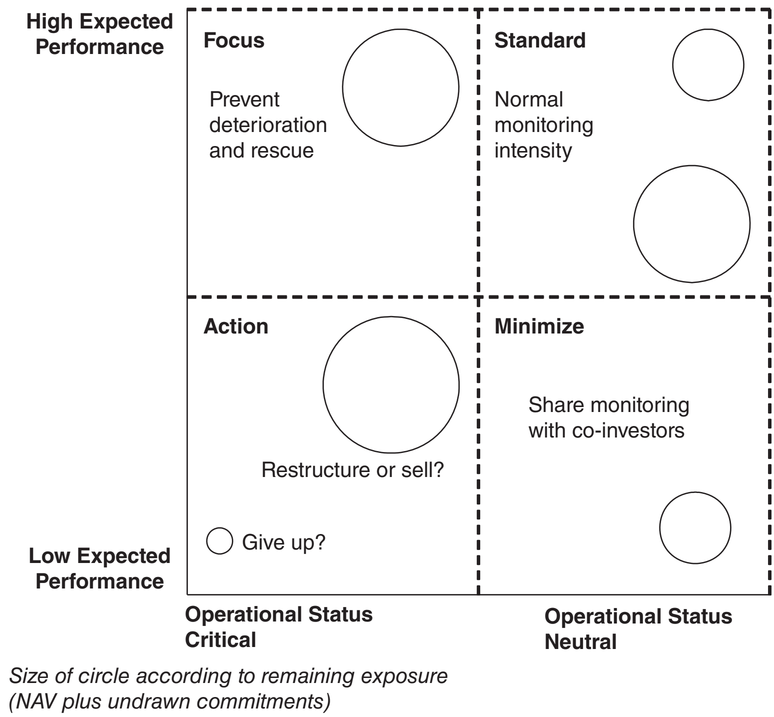


EXHIBIT 9.10 Monitoring Intensity Management

and no credible alternative is brought forward, investors can influence the fund manager to reduce management fees or even release LPs from portions of their commitments. This often means taking part in the decision-making process, vetoing decisions,[21] or simply exercising pressure on the fund manager. More often than not, GPs give in to reducing fund size. This investor-friendly behavior can build up goodwill and ease the next fundraising exercise.

2. One of the more obvious and common monitoring actions is the increased use of renegotiation of management fees and fund size toward the end of a fund's life.
3. Of course, GPs have the right to refuse such voluntary actions. However, this often results in pressure, activism, and even lawsuits from LPs who want some or all of their money back. In the extreme, and if there is an agreement among LPs, the fund management team can be terminated without cause. Even without recourse to such extreme measures, the threat of action or the noise of complaints from investors can be highly damaging to the reputation of a fund manager. This in turn can have serious implications for future fundraising ambitions.

9.7.2 Two Actions outside the Fund's Governance Process

The two actions outside the fund's governance process are:

1. In situations in which a fund management team has clearly demonstrated that it is not up to the job or that it is not cooperating with its LPs, the simplest and

most obvious action of an LP is not to commit to the follow-on fund. This is also most feared by the fund manager, as often the loss of a reputable investor sends a clear negative signal to the market. Not only would the team need to go back to the capital market for fundraising, but it would also do so with a tarnished reputation.

2. Investor default is questionable, as it constitutes a contractual breach, but this refusal to pay capital calls when requested may be the instrument of last resort if the fund manager is clearly incompetent.[22] There can be substantial penalties when an LP does not pay a requested drawdown. The LP would likely forfeit all prior investments in the fund. The investor might also find it difficult to continue investing in private equity, as other fund managers might refuse to accept new commitments from investors who did not fulfill their prior obligations.

9.7.3 Two Actions outside Fund Investing

The two actions outside fund investing are:

1. Co-investing alongside fund managers is a tool for increasing exposure to certain sectors, but it also requires specific skills.
2. Another option is to adjust allocations using public small-cap equities as a proxy for PE investments. This could be a meaningful investment in which a significant pool of undrawn commitments needs to be managed.

Due to the illiquidity of PE funds, the tool set for the management of the portfolio is comparatively restricted. One such tool is the active trading of positions through the secondary market. At the fund level, the main approach to adjusting the portfolio structure is the continuous review of the ongoing investment pipeline in primary fund investments. However, this strategy may be restricted by the availability of suitable investment opportunities or by delays in closing the deals. Moreover, when the PE portfolio is large, the adjustment via primary transactions may be slow and insufficient. With some restrictions, as discussed in the following section, the secondary market offers an ability to more quickly adjust exposures in the private equity portfolio.[23]

9.8 THE SECONDARY MARKET

Secondary transactions refer to the buying and selling of preexisting limited partnership interests in private equity and other alternative investment funds. Kießlich (2004) suggests that the first secondary PE transaction took place in 1979, with David Carr buying Thomas J. Watson's stake in a VC fund when Watson had to exit after being appointed the new ambassador to the Soviet Union. The genesis of the secondary market is said to date back to the infamous Black Monday in October 1987 and to the world economic crisis of the early 1990s. These two events produced a large need for liquidity among many financial institutions, especially those with illiquid assets such as private equity, which in turn created a new market for secondary interests in PE funds or companies.

Since then, a sizable secondary market has emerged, which allows investors to sell their interests in limited partnerships in order to generate liquidity or pursue strategic objectives.

9.8.1 Market Development

The development of a secondary market has usually been portrayed as a market response to the illiquidity of fund investments that constrains the universe of investors to those with appropriate liability profiles. Often intermediated by specialized investment banks, secondary stakes are bought by investors who are primarily attracted by the potential discount at which a transaction may take place, the shorter period during which the invested capital is locked in, and the portfolio diversification properties of secondaries.

Historically, the secondary market has been largely driven by the sale of interests in PE funds, with LPs being forced to exit prematurely. Initially, the secondary market was viewed by many as a market of last resort for those desperate to liquidate PE positions. It not only suffered from a limited track record but also largely lacked transparency and was characterized by an unhealthy predominance of buyers over sellers. A number of recent developments have led to changes in this perception:

- Large institutions regularly exit from private equity as part of an overall portfolio management and reallocation strategy.
- Some LPs entering the PE market are interested in secondary transactions as a means of getting *exposure to PE assets quickly* without having to commit capital over a period of 10 to 12 years, as is typical for primary fund investors. Given that it normally takes several years for a fund to reach the cash flow break-even point, a secondary acquisition of a four-year-old fund reduces this period significantly and potentially even eliminates it.
- Such LPs also seek a *reduction of risks* by using secondary transactions in order to achieve a wider and more balanced vintage-year spread.
- While the secondaries' total value to paid-in (TVPI) ratios are generally lower due to their shorter life, the *strong IRR performance* has significantly contributed to the rising amount of commitments to secondary funds.
- The different dynamic in the secondary market can be employed as a means to *counteract the J-curve effect* of primary fund investments.
- There is increasing experience in conducting secondary transactions that are considered successful by buyers as well as sellers.
- Respected players, such as U.S. university endowments and pension funds, are routinely investing in secondary funds-of-funds specialists.

Today, investing in secondary transactions has become an accepted investment strategy and an active portfolio management tool in its own right, and the market has grown substantially in size and maturity. Additionally, in recent years an active market for the sale of directly held portfolio companies has emerged. Here, it is the GPs who want or are under pressure to sell, mainly as their fund is approaching the end of its contractual lifetime. In such **synthetic secondaries**, portfolio companies are packaged up and sold to another manager, usually with the backing of a secondary fund specialist. In fact, dedicated secondary direct strategies have only recently become recognized as independent strategies requiring highly specialized fund managers. This market is attractive due to its inefficiencies, but such opportunities are hard to value, and the due diligence is complex.

9.8.2 Market Size

The secondary market began to take off following the burst of the dot-com bubble. At that time, institutions and particularly individual investors who had previously enthusiastically committed to VC funds felt the liquidity squeeze acutely and often could not afford to meet further capital calls.[24]

It is estimated that the global volume of secondary transactions increased tenfold in the first decade of the 21st century.[25] This market has grown significantly, mirroring the substantial expansion in the primary fundraising market in private equity. Secondaries were probably about 10% of the total PE market in 2005. It is important to emphasize that the total supply was significantly larger, however.

Recent data by Cogent Partners (2015) show a steady increase in the volume of transactions in the secondary market: from \$25 billion in 2011 to \$42 billion in 2014. Also, as shown in Exhibit 9.11, the secondary transactions have been taking place close to the NAV in recent years. These transactions took place at discounts of about 40% in 2009 and about 15% in 2010–12. This points out that the secondary market is no longer driven entirely by the desire of LPs to sell their positions because of liquidity needs.

Despite its increasing importance and growth in the volume of transactions, the secondary market has remained highly opaque. Although various data vendors tried to build specific Internet platforms for secondary transactions, information about market conditions is often spurious, and little is known about prices at which transactions finally settle. Historically, around 3% to 5% of outstanding LP exposure comes to market, which represents the pool from which potential buyers can fish. In the end, less than 2% of the outstanding exposure, calculated as the sum of NAV and unfunded commitments, is actually traded each year.[26]

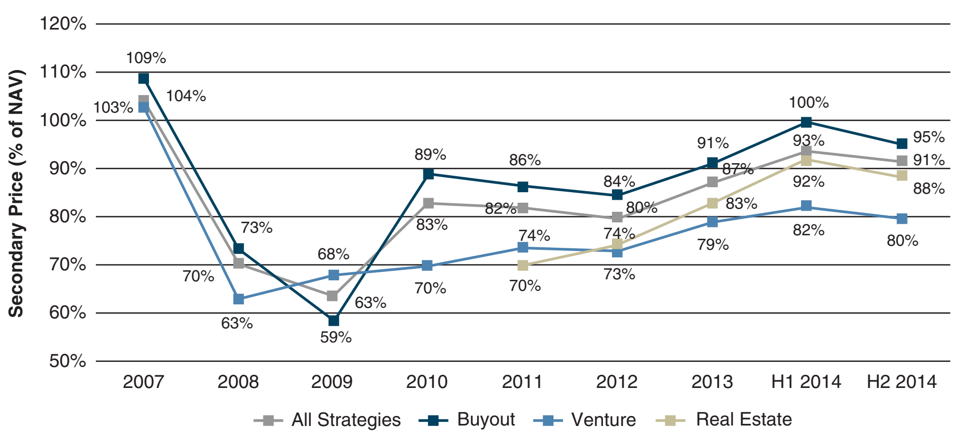


EXHIBIT 9.11 Evolution of Average Secondary Market Pricing
Source: Cogent Partners (2015).

9.8.3 Buyer Motivations

On the buy side, the secondary market has historically been dominated by dedicated secondary funds and other funds of funds. Apart from the portfolio management considerations mentioned previously, the obvious reasons why investors want to buy relate mainly to the potential attractiveness of secondary stakes.

- To begin with, secondaries are often distressed sales that take place at a significant discount to the most recent NAV. Under certain market conditions, stakes in funds are offered by LPs who are generally in need of liquidity or, specifically related to their PE activities, face a nontrivial risk of default on future capital calls from fund managers. Such assets can be a bargain for those who have at their disposal a sufficient degree of liquidity to meet the unfunded commitments of the fund investments.
- An important advantage of secondary investments is that they are subject to less uncertainty when compared with commitments to primary funds, which represent blind pools of capital. At the time of making a commitment, a primary investor does not know how his capital will eventually be deployed by the fund—apart from the broad investment guidelines specified in the limited partnership agreement. By the time a secondary transaction takes place, a significant share of the fund's capital has already been invested in portfolio companies. Prospective buyers can analyze them in detail and have indicators that help distinguish between companies that developed according to plan and those that did not.[27]

In recent years, therefore, a growing number of nontraditional buyers have entered the market, including pension funds, insurance companies, endowments, foundations, family offices, sovereign wealth funds, and hedge funds.

9.8.4 Seller Motivations

The reasons why LPs decide to dispose of their holdings on the secondary market are as diverse as those for buyers. Broadly speaking, one may differentiate between sales that are motivated primarily by liquidity constraints and those that are related to the strategic repositioning of an investor's portfolio. Understanding these different reasons is important for assessing the informational content of transaction prices in the secondary market.

- Sellers may be motivated by active portfolio management considerations. As the secondary market has matured, it has been increasingly used strategically. Sell-side LPs may decide to divest their entire stake in a particular fund or even offer their entire portfolio of fund investments. Others, however, may decide to sell only a certain share of their stake in a partnership, signaling their continued confidence in the fund manager.
- Sometimes LPs are dissatisfied with the performance of a fund or a portfolio of funds and lose confidence in a management team. Selling their stakes in such funds can generate capital that can be redeployed into new opportunities.
- Conversely, an LP may want to lock in returns in situations in which the fund manager is not likely to materially increase the performance of the underlying portfolio companies during the remainder of the lifetime of the fund.

- Another strategic reason to sell may lie in an unintended overexposure to private equity. This so-called denominator effect usually arises in periods of financial stress, when prices of marketable instruments adjust much faster than valuations of illiquid investments, resulting in a higher-than-targeted share of the latter.

As far as liquidity-motivated sales are concerned, the desire to transfer stakes in a limited partnership is usually cyclical, in line with macroeconomic developments. Despite such good reasons for an LP to sell, this raises the question of whether an early exit is against the main idea of private equity, namely to be patient until the "pearls" emerge, and, by and large, institutions are justifiably reluctant to sell.[28] Although the existing investor may occasionally be under pressure to exit a fund, from an economic point of view the case for selling is far from clear under normal circumstances and raises the question of sourcing and proper pricing of secondary transactions.

9.8.5 Secondary Market Investment Process

Secondary transactions can be difficult to access and tend to be open only to well-connected investors who are trusted to efficiently assess PE assets, safeguard confidential information, and execute the transaction with minimal fuss. Conversely, the secondary sales process can be highly complex and time-consuming.

9.8.6 Sourcing Secondary Opportunities

While transactions involving individual funds or small portfolios are usually negotiated bilaterally, intermediated auctions have become increasingly common for larger and more complex transactions. This process is facilitated by financial intermediaries, who are today involved in the majority of deals in the secondary market and ensure that the sales process is competitive. Used mainly by sellers, intermediaries help identify buyers and structure offerings of fund interests. Financial intermediaries are either specialized secondary advisers or placement agents, generally charging a transaction fee between 1% and 2% of the value of the transaction.[29]

Typically, a number of LPs are contacted to sound out their potential interest. In cases in which a large portfolio is being sold, intermediaries may divide it into subsets to improve the chances of finding interested buyers. LPs make bids in a managed auction for the particular stakes they want to purchase; these decisions are based on confidential information the intermediary provides about the funds' holdings and their valuation. Many secondary auctions involve two rounds. After a first round of bidding, the seller and the intermediary invite a subgroup of potential buyers to participate in the second round. During this round, interested buyers have the opportunity to revise their bids in light of new information they might have acquired in the process. While the set of information the intermediary provides is the same for all interested buyers, large LPs with diversified portfolios often have an important competitive advantage in that they have superior proprietary information about the fund manager, the manager's previous track record, and the quality of the current investments.

9.8.7 Pricing Secondary Stakes

While prices in the secondary market are generally expressed in terms of discounts or premiums relative to a fund's NAV, this can be misleading. To determine a fund's fair value, the NAV reported by the GP is of limited relevance. Valuations of funds can be considered to be fair only if they are equal to the present value of the fund's overall expected cash flows, which include not only those that are related to the investments a fund has already made but also future cash flows that are associated with the undrawn commitments the buyer needs to fund.

The secondary price is estimated by discounting the expected future cash flows, based on the buyer's expectations about the asset's performance:

$$P_0 = \sum_{t=0}^{T} \frac{CF_t}{(1 + IRR_{Buyer})^t} \tag{9.1}$$

with P_0 denoting the secondary price offered by the buyer, CF_i the fund's expected cash flow at time i, n the fund's maturity, and IRR_{Buyer} the buyer's required or expected IRR. This price reflects the expected cash flows from both funded and unfunded commitments and thus will generally be different from the NAV reported by the GP; this is because the accounting impact is usually the reference point for the seller. The discount can be expressed as follows:

$$\text{Discount}_t = \frac{NAV_t - P_t}{NAV_t} \tag{9.2}$$

APPLICATION 9.8.7A

A prospective secondary-market private equity investor forecasts a total of five equal cash flows from a fund of \$1 million per year over the next five years (i.e., years 1–5). The buyer's required internal rate of return (IRR) is 18%. The fund's current net asset value (NAV) is \$3.5 million. What is the fund's estimated discount?

The present value of the cash flow stream at 18% is \$3.127 million using Equation 9.1. The discount using Equation 9.2 is [(3.500 – 3.127)/3.500)], which equals 10.7%.

The cash flow projection (see Chapter 12) needs to take a number of factors into account: in particular, the expected exit value and exit timing for current portfolio investments, projected future capital calls, and the return on future investments made using such drawdowns. **Exit value** refers to the price that the fund can receive when portfolio companies are sold through initial public offerings (IPOs) or strategic sales. **Exit timing** refers to the period during which portfolio companies are expected to be sold and exit values are realized. This analysis is generally based on a combination of GP guidance, co-investor insight, and market analysis. All public investments are valued using a mark-to-market analysis, taking potential lockups into consideration.

The resulting projected gross cash flows are then run through the partnership's legal structure, taking future management fees and the distribution waterfall into account to arrive at a net cash flow stream for the fund. Once the projected net cash flows are determined, the present value is calculated by discounting the net cash flows at an appropriate discount rate (see Chapter 13). This target rate varies across market segments, with mezzanine funds usually having a lower target rate than buyout funds, which in turn have a lower target rate than VC funds, reflecting their specific risk characteristics. Furthermore, target rates mirror market conditions, which vary over time.[30]

Secondary market prices are only of limited use as benchmarks for fair market value, as they are not observable. Transactions are confidential, and the final settlement price is generally known to only the buyer and the seller (and the intermediary to the extent that a transaction has been facilitated by a specialist agent). The secondary market price reflects current market conditions for those who participate in the market as either sellers or buyers but reveals little about the underlying value of the portfolio, which is held until maturity.

Exhibit 9.11 displays secondary market pricing history, which is expressed as average high bids as a percentage of NAV. First, we can see that the secondary prices are almost always discounted relative to NAV. Second, the discounts reached their highest values during the most recent financial crisis. Third, the average discounts for buyout funds are typically smaller than those of venture capital funds. This can be explained by the fact that portfolio companies of venture capital funds are more difficult to price, and therefore there is greater uncertainty regarding their NAVs.

9.8.8 Limitations of the Secondary Market

Undoubtedly, the secondary market would not have reached its current level of significance if it had not brought about important advantages for both sellers and buyers of interests in limited partnerships. However, as the secondary market has not materially altered the fundamental characteristics of fund investments, the challenges investors face in measuring and managing the particular risks that come with commitments to limited partnerships remain essentially unchanged. As significant as the development of the secondary market may be for constructing efficient portfolios, it should not be perceived as a game changer in terms of risk management in illiquid investments.

Although the secondary market has expanded rapidly in recent years, it has remained small relative to the primary market, with only a few percent of primary commitments being transacted in the secondary market. Importantly, the secondary market, just like other financial markets, may dry up precisely at the time when it is needed most. The year 2009 provides a warning. Although a maximum amount of supply came to market, the actual transaction volume collapsed, as sellers were not prepared to accept discounts that were at times 60% or higher unless they were under exceptional pressure to create liquidity.

9.9 CONCLUSION

Manager selection is crucial for alternative investments, as there are typically significant performance differences between top-quartile and bottom-quartile managers. This is different from asset managers in the traditional space, where such differences are rather small. This means that while the marginal benefits of spending significant resources on selecting top-quartile traditional asset managers might be small, private equity investors will benefit tremendously from spending significantly on manager selection.

There are two issues related to selecting top-quartile managers. First, and more important, is whether ex ante one can identify top-quartile managers. Again, while research has shown that there is little persistence among traditional asset managers, evidence supports the notion that alternative investment managers tend to display performance persistence, and this appears to be strongest among PE managers.

Second is the crucial question of gaining access to top-quartile managers. Since PE funds have limited capacity, they cannot accommodate all potential investors. As a result, newcomers to the PE investing industry may have difficulty accessing top-quartile managers. One possible solution to this problem is to use funds of funds, which are more likely to have access to top-quartile managers than are investors who are just getting started.

NOTES

1. See Chung (2012); Braun, Jenkinson, and Stoff (2013); and Harris, Jenkinson, and Stucke (2014).
2. See, for instance, Von Braun (2000), Tierney and Folkerts-Landau (2001), Scardino (2004), Kaplan and Schoar (2005), Rouvinez (2006), Hendershott (2007), Sørensen (2007), and Phalippou and Gottschalg (2009).
3. Meerkatt et al. (2008).
4. Talmor and Vasvari (2011).
5. Harris, Jenkinson, and Stucke (2014).
6. Korteweg and Sørensen (2014).
7. Chung (2012).
8. Korteweg and Sørensen (2014) concluded that "top-quartile performance does not necessarily imply top-quartile skills," making it difficult for investors to differentiate between luck and skill. Some LPs try to do the analysis on the underlying portfolio company data, where the statistical significance will be reached more quickly. This may give information on many aspects of the GP's investment approach but not on the GP's portfolio management.
9. Conner (2005).
10. Kaplan and Schoar (2005).
11. This is consistent with Manninena, Jääskeläinena, and Maula (2010), who found no significant institutional learning among GPs.
12. See Cornelius, Juttmann, and de Veer (2009). Chung (2012) comes to similar conclusions: that persistence is largely explained by common market conditions facing funds raised in succession, and that this result does not support the view that GPs have differential and proprietary skills.
13. See, for instance, Carhart (1997) and Soe and Luo (2012).
14. In fact, there has long been an argument, not to mention a sprawling literature (Tom Peters and Robert Waterman's classic *In Search of Excellence*, from 1982, and Jim Collins's *Good to Great*, published in 2001, for instance) that some companies are truly better and will outperform through good times and bad times. But taken together and viewed over a longer period, the story these "excellent company" books tell is not one of sustainable competitive advantage and enduring high performance. See Beinhocker (2006); Rosenzweig (2007); and Raynor, Ahmed, and Henderson (2009).
15. See Burgel (2000) or Conner (2005).

16. In the set of data used by Conner (2005), firms that raised a subsequent fund did so 2.9 years on average after raising the predecessor fund.
17. See also Lerner and Schoar (2004). The authors presented the theory that by choosing the degree of illiquidity of the security, private equity fund managers can influence the type of investor the firm will end up attracting. This allows managers to screen for deep-pocketed investors (i.e., those who have a low likelihood of facing a liquidity shock), as they can reduce the GP's cost of capital in future fundraising efforts. The authors' analysis is based on the assumption of an information asymmetry about the quality of the manager between the existing investors and the market. The GP faces a problem when having to raise funds for a subsequent fund from outside investors, because the outsiders cannot determine whether the manager is of poor quality or the existing investors were hit by a liquidity shock. Transferability constraints are less prevalent when private equity funds have LPs that are known to be subject to few liquidity shocks (e.g., endowments, foundations, and other investors with long-term commitments to private equity).
18. Being perceived as a professional and serious investor also increases negotiation power vis-à-vis the fund managers. In a comparatively small industry, a strong network is a credible threat against a team that would otherwise be unwilling to compromise.
19. Also, the administration of capital calls and distributions by GPs offer opportunities to conduct monitoring by LPs, as in this context, for example, contractual terms can and should be checked.
20. We expect that information targeting downside protection will be more likely to be standardized and shared among all LPs. Ultimately, to remove the fund manager, a majority vote of the LPs is required. Therefore, it is in all investors' interests that everybody knows how the portfolio is performing.
21. An illustrative list of considerations would be the appointment and remuneration of managers, the approval of annual reports, the approval of the budget, decisions on unforeseen investments (loans or side funds), issues of new shares, extension of the investment period, and so on.
22. See Meek (2003): "In one instance the limited partners of a U.S. fund have simply refused to honor any future drawdowns, taking the view that to do so would simply be throwing good money after bad."
23. There are other, albeit technically more complicated ways of exiting investments in PE funds before maturity. For instance, securitization involves the transfer of the limited partnership share to a special purpose vehicle (SPV; a special entity, typically located in a tax-efficient or legally efficient offshore location, established by a company to answer a certain financial or legal problem, e.g., to pay lower taxes) for a so-called collateralized fund obligation. The SPV is a distinct legal entity that issues senior and junior notes and uses the proceeds from the issuance to invest in a PE fund of funds. For such securitizations, see Chapter 19 in Mathonet and Meyer (2007).
24. Brown and Berman (2003) reported that in the United States at the time, 400 individuals were looking for bailouts and were willing to sell at a discount rather than become defaulting investors. Secondary specialists like Coller Capital or Landmark Partners were even able to close a number of so-called walk-away deals in which they paid the original investors nothing for the position and in exchange agreed to cover future capital calls.
25. See Cornelius et al. (2013).
26. See Cornelius et al. (2013).
27. Note, however, that this advantage can be significantly reduced, and even eliminated, in early secondary transactions. Sometimes called purchased primaries, such transactions take place at a very early stage of the fund. In extreme cases, the buyer agrees to buy the seller's commitment at a time when the fund has not yet made any acquisitions.
28. For example, according to Private Equity International (2006), for CalPERS, conducting a sale is not an option.

29. See Talmor and Vasvari (2011).
30. See Cornelius et al. (2013).

REFERENCES

Beinhocker, E. D. 2006. *The Origins of Wealth*. London: Random House.

Braun, R., T. Jenkinson, and I. Stoff. 2013. "How Persistent Is Private Equity Performance? Evidence from Deal-Level Data." Working paper, August 13. Available at http://ssrn.com/abstract=2314400 or http://dx.doi.org/10.2139/ssrn.2314400.

Brown, E., and P. Berman. 2003. "Take My Venture Fund—Please!" *Forbes*, June 23.

Burgel, O. 2000. *UK Venture Capital and Venture Capital as an Asset Class for Institutional Investors*. London: BVCA.

Carhart, M. M. 1997. "On Persistence of Mutual Funds Performance." *Journal of Finance* 52 (1).

Chung, J.-W. 2012. "Performance Persistence in Private Equity Funds." Working paper, Chinese University of Hong Kong, November 24.

Cogent Partners. 2015. "Secondary Market Trends and Outlook." White paper, January.

Conner, A. 2005. "Persistence in Venture Capital Returns." Private Equity International, March.

Cornelius, P., C. Diller, D. Guennoc, and T. Meyer. 2013. *Mastering Illiquidity: Risk Management for Portfolios of Limited Partnership Funds*. Chichester, UK: John Wiley & Sons.

Cornelius, P., K. Juttmann, and R. de Veer. 2009. "Industry Cycles and the Performance of Buyout Funds." *Journal of Private Equity* 12 (4): 14–21.

Cullen, A. 2004. "Locating Venture Returns." Working Knowledge, Harvard Business School, February 16.

Cumming, D., G. Fleming, and A. Schwienbacher. 2004. "Style Drift in Private Equity." Working paper, Center for International Securities and Derivatives Markets, May.

Harris, R. S., T. Jenkinson, S.N. Kaplan and R. Stucke. 2014. "Has Persistence Persisted in Private Equity? Evidence from Buyout and Venture Capital Funds." February 28. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2304808.

Hendershott, R. 2007. "Using Past Performance to Infer Investment Manager Ability." Preliminary paper, November. Available at www.scu.edu/business/mindwork/winter08/upload/hendershott-past_performance_nov07.pdf.

Kaplan, S. N., and A. Schoar. 2005. "Private Equity Performance: Returns, Persistence, and Capital Flows." *Journal of Finance* 60.

Kießlich, U. P. 2004. "The Secondary Market for Private Equity Funds." Diploma thesis, Wirtschaftliche Hochschule fűr Unternehmensfűhrung, Vallendar, Germany, August 2.

Korteweg, A. G., and M. Sørensen. 2014. "Skill and Luck in Private Equity Performance." Working Paper 179, Rock Center for Corporate Governance at Stanford University, October 29. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2419299.

Lerner, J., and A. Schoar. 2004. "The Illiquidity Puzzle: Theory and Evidence from Private Equity." *Journal of Financial Economics* 72 (1): 3–40.

Lerner, J., A. Schoar, and W. Wongsunwai. 2007. "Smart Institutions, Foolish Choices? The Limited Partner Performance Puzzle." *Journal of Finance* 62 (2): 731–64.

Manninena, O., M. Jääskeläinena, and M. Maula. 2010. "Access versus Selection: What Drives Limited Partners' Private Equity Returns?" Working paper, Aalto University, Finland, November 1.

Marquez, R., V. Nanda, and M. Yavuz. 2010. "Private Equity Fund Returns: Do Managers Actually Leave Money on the Table?" Working paper, Washington University, St. Louis, March 17.

Mathonet, P.-Y., and T. Meyer. 2007. *J-Curve Exposure: Managing a Portfolio of Venture Capital and Private Equity Funds*. Chichester, UK: John Wiley & Sons.

Meek, V. 2003. *Time to Deviate from the Standard?* London: AltAssets.

Meerkatt, H., J. Rose, M. Brigl, H. Liechtenstein, M. J. Prats, and A. Herrera. 2008. *The Advantage of Persistence: How the Best Private-Equity Firms "Beat the Fade."* Boston: Boston Consulting Group and IESE Business School of the University of Navarra. www.bcg.com/documents/file15196.pdf.

Phalippou, L., and O. Gottschalg. 2009. "The Performance of Private Equity Funds." *Review of Financial Studies* 22 (4): 1747–76.

Private Equity International. 2006. "Orphans in the Portfolio." May.

Raynor, M. E., M. Ahmed, and A. D. Henderson. 2009. "A Random Search for Excellence: Why 'Great Company' Research Delivers Fables and Not Facts." Deloitte Development, December.

Robbie, K., M. Wright, and B. Chiplin. 1997. "The Monitoring of Venture Capital Firms." *Entrepreneurship Theory and Practice* 21 (4): 9–28.

Rosenzweig, P. 2007. *The Halo Effect and the Eight Other Delusions That Deceive Managers*. New York: Free Press.

Rouvinez, C. 2006. "Top Quartile Persistence in Private Equity." Private Equity International, June.

Scardino, J. 2004. "Past Performance a Guide to Likely Future Performance in Private Equity." Private Equity Monitor.

Soe, A. M., and F. Luo. 2012. "Does Past Performance Matter? S&P Persistence Scorecard." Standard & Poor's Research, June 7. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2079822.

Sørensen, M. 2007. "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital." *Journal of Finance* 62.

Talmor, E., and F. Vasvari. 2011. "Gatekeepers." In *International Private Equity*. Chichester, UK: John Wiley & Sons.

Tierney, J. F., and D. Folkerts-Landau. 2001. *Structured Private Equity: An Old Market Becomes an Emerging Asset Class*. Frankfurt: Deutsche Bank Global Markets Research.

Von Braun, E. 2000. "Selektion und Strukturierung von Private Equity Fonds-Portfolios." Paper presented at the IIR Private Equity & Venture Capital Conference, Munich.

CHAPTER 10

Private Equity Operational Due Diligence

This chapter provides an overview of the operational due diligence process for investors in private equity (PE) funds. The chapter begins by outlining common operational risk areas covered during an operational due diligence review. It then goes on to outline the importance of private equity operational due diligence and provide an overview of the operational due diligence process from initial assessment to ongoing monitoring. Finally, it concludes by discussing the potential benefits for performing ongoing operational due diligence.[1]

10.1 THE SCOPE AND IMPORTANCE OF OPERATIONAL DUE DILIGENCE

This section provides an introduction to operational due diligence and its relationship to investment due diligence.

10.1.1 Operational Due Diligence Is Driven by Operational Risk

Operational due diligence (ODD) refers to the process of evaluating operational risk—a specific type of risk that must be addressed to ensure that investors will not be subject to financial or reputational risks of investing in funds that may experience large losses for noninvestment reasons. *Operational risk* is a general term that refers to a series of risks that are not purely investment related in nature. But what exactly are these risks? Exhibit 10.1 provides a summary of key operational risk areas typically covered by investors during ODD reviews to ensure that risks are appropriately addressed by the manager.

Historically, ODD reviews by investors were primarily conducted with a limited scope involving noninvestment-related aspects of fund operations, such as fund accounting. Over time, ODD has evolved beyond traditional back office operations in terms of both the scope and depth of the operational risk areas examined.

EXHIBIT 10.1 Key Operational Risk Areas Typically Reviewed during Operational Due Diligence

Postinvestment operational processes	Cash management
Valuation	Collateral and custody management
Legal function	Compliance function
Fund and firm service providers	Fund reporting
Legal documentation	Financial documentation
Meta risks	Information technology
Information security	Business continuity and disaster recovery
Board of directors	Tax practices
Insurance	Counterparty management
Human capital	Assets under management and fund flows

10.1.2 Distinguishing Operational Due Diligence from Investment Due Diligence

In its most basic form, investor due diligence on fund managers, be they hedge funds, private equity (PE), real estate funds, or traditional long-only managers, can be divided into two main areas.

The first area, **investment due diligence (IDD)**, focuses on gaining a detailed understanding of a fund manager's investment strategy with the ultimate goal of determining whether or not the strategy is appropriate for the investor. The areas covered by IDD include markets traded, use of leverage and derivatives, anticipated net market exposure, net market exposure limits, qualifications of the investment team, and risk management procedures. In developing a complete risk assessment of a private equity fund, it is important to analyze not only investment-related risks but also operational risks.

The second area, operational due diligence, focuses on all aspects of fund managers other than those that are purely investment related. To clarify, this does not mean that the specification of risk factors covered during an ODD review is intended to preclude an investor from considering the investment aspects of a fund. In fact, as you will see further along in our discussion, maintaining a thorough understanding of a fund's investment strategy can be critical to driving the ODD process.

10.1.3 Why Operational Due Diligence Is So Important in Private Equity

It is especially important to evaluate operational risks in PE funds due to the longer-term investment horizon traditionally associated with PE investing, typically reflected through the locking up of capital for periods of several years. This can be contrasted with hedge fund investing, which may provide investors with the ability to redeem capital over a much shorter time frame compared to PE investments. Longer investment periods increase the probability and magnitude of exposures to operational risks. With limited liquidity and therefore limited remedies, private equity investors are especially well advised to perform significant operational due diligence prior to committing capital for these long periods of time to ensure that they fully vet the investment and operational risks associated with a fund.

10.1.4 Five Key Benefits of Operational Due Diligence for Private Equity Funds

Five key benefits to performing ODD on private equity funds are:

1. Enhanced understanding of a fund's back office operational processes and procedures
2. Diagnosis of how scalable PE fund operations are from a personnel and systems perspective to facilitate growth, including future fund launches
3. Analysis of the ability of a firm to adapt to future exogenous operational factors, including changes in regulatory rules and new accounting pronouncements
4. Accumulation of useful information regarding risk modeling and management
5. Diagnostic benefits to limited partners (LPs), including further understanding of a fund's mechanics (i.e., operational details) across a number of operational risk areas, ranging from information technology to cash management

10.1.5 Four Explanations for the Expanding Scope of Operational Due Diligence

There are a number of historical explanations for the expanding scope of items covered during operational due diligence. These include the following:

1. Increased availability of overall due diligence resources. As investing in alternatives has become more prevalent, investors are devoting more time and resources toward due diligence on alternative funds, which has included a focus on ODD.
2. Continued presence of fund fraud. The occurrence of a number of fund failures across the alternatives space has led to a driving interest in ODD. Unfortunately, many of these fund failures have been outright frauds, often with operational reasons at their center. These incidents have served as a wake-up call to many investors by alerting them to the fact that they had been underinvestigating operational risks.
3. Increasing operational sophistication. There has been a growing sophistication in the complexity of alternative investment fund operations. This increased complexity has facilitated the need for more in-depth analysis among investors in an attempt to ensure that alternative investment managers are appropriately navigating this increased complexity.
4. Expanded regulatory compliance. There has been a steady stream of fund regulations, such as the Dodd-Frank Act in the United States and the Alternative Investment Fund Managers Directive (AIFMD) in Europe, which have raised the overall complexity of various regulatory areas, including compliance management. Analysis of such areas has traditionally fallen within the scope of ODD; thus, by association, the scope of ODD has expanded to comport with this increased growth.

10.2 EIGHT CORE ELEMENTS OF THE OPERATIONAL DUE DILIGENCE PROCESS

The alternative investment ODD process employed by investors has seven primary steps, as outlined in Exhibit 10.2, and an eighth step (ongoing monitoring) in the event that an investment is made.

The core elements of this process are the same for various asset classes. For example, operational due diligence on both a hedge fund and a private equity fund require document collection and document analysis. While there are commonalities in due diligence processes across hedge funds and private equity, each asset class and specific strategy will have detailed questions and processes that differ from those needed in other investment areas.

In this chapter, we will walk through steps one through four and step eight while including reference to unique considerations of private equity funds. Due to their relatively universal nature, steps five through seven are covered in the accompanying chapter on hedge fund operational due diligence (Chapter 33).

Before beginning this discussion, we should clarify a few general things about these steps. When conducting an ODD review, these steps are not always carried out

EXHIBIT 10.2 The Operational Due Diligence Process

in a sequential order. For example, service provider review and confirmation could take place in conjunction with the document collection process. In practice, this is often the case. In order to promote process efficiency, conducting different stages of the ODD process in parallel can expedite the process.

Readers should be cautious, however, to note that speed of ODD process review should not be sought at the expense of quality. Of course, you could conduct a cursory service provider review before the on-site visit with the manager, but as an investor, you might glean useful insights from a fund manager during the on-site visit that could be used to facilitate a more detailed service provider review. Therefore, in order to maximize the review effort, what is presented here is the general order in which the process typically takes place.

It is also worth noting that just because an LP may feel that a stage of the review process has been completed, it does not mean that investors are not allowed to go back and revisit any one stage. For example, let's say that the LPs have gone through an initial round of document collection from a manager. Suppose that, when reading the documents, they discover that certain other documents were referenced that they didn't think to ask for initially. Does this mean that the investors have missed their chance? Of course not.

An investor performing ODD will often go back to the manager several times requesting additional documents. It is worth nothing, however, that there are overriding practical considerations that must be taken into account. In order to facilitate an efficient ODD process, LPs must be conscious of the time needed by general partners (GPs) to respond to requests. Thus, LPs must engage in a balancing act so as not to overly burden the fund manager with too many separate requests. Instead, it is often considered prudent to bundle requests for follow-up items, such as additional documents. Now let us proceed with our discussion of the ODD process.

10.3 PRIVATE EQUITY OPERATIONAL DUE DILIGENCE DOCUMENT COLLECTION PROCESS

One of the primary purposes of document collection is to allow investors to begin to develop an initial assessment of a fund manager's operational risk profile before the on-site visit. An **operational risk profile** is an outline or summary of potential losses or other exposures of a fund due to errors or failures within the fund's functions other than those purely attributable to the fund's investment strategy.

To clarify, there are three primary sources of documentation for investors during an ODD review:

1. The first source is directly from fund managers themselves. These are the types of documents that will be referenced in this section.
2. Fund service providers are a second common source from which ODD documents may be collected. We address these types of documents in the service provider section.
3. The third source is documentation external to the GP and service provider relationship, such as regulatory filings.

What documents should investors collect? Exhibit 10.3 outlines a list of key documents commonly collected from private equity GPs during an ODD review.

There is no universal set of terms that private equity fund managers use when naming documents. One example is the compliance manual. Some private equity managers may maintain a document similar to a compliance manual but refer to it as a code of ethics. It is good to be aware of this, as some GPs are more literal than others in responding to LP requests.

If an investor does not refer to the exact name used by the fund when requesting a certain document, the GP might simply reply that the fund does not maintain the document requested. Continuing the previous example, the code of ethics document may serve an investor's goal at this stage of collecting the firm's compliance manual; however, the fund may not provide it because the investor didn't specifically ask for it by name, requesting a compliance manual instead.

Oftentimes, this problem can be overcome by LPs in two ways. The first is by engaging in discussions with GPs when submitting document requests to provide perspective on the goals of the LP document request. Practical experience suggests that this will commonly lead to more responsive and detailed document request and collection procedures. A second common way is to develop documentation requests in such a way as to not be overly self-limiting.

Continuing our example, instead of asking for just the compliance manual, an LP could submit a request for "the compliance manual and any other related compliance policies and procedures." This broader request would likely put the fund on notice that the LP was interested in collecting the compliance manual, the code of ethics, and any other relevant compliance-related documentation the fund maintains.

Another common issue that typically arises during the LP document collection stage is that a fund manager may choose not to distribute certain documentation outside the fund's office. There are a myriad of reasons GPs typically cite for this, including confidentiality concerns. In this case, there are two common approaches LPs typically employ.

The first is to collect what is known as **compromise documentation**, which allows investors to collect part of a document or a sample of a document for their files while still allowing the fund to appropriately manage any document distribution concerns.

One common document around which such conflicts arise is the compliance manual. Some GPs view the contents of this document as proprietary and, as such, do not like to distribute it to investors outside the office. One way that many LPs negotiate compromise documentation relating to the compliance manual is to request the table of contents of the manual. This table of contents not only serves as a placeholder in an LP's due diligence files with regard to compliance-related documentation but also provides the LP with an understanding of what items are covered within the compliance manual itself, therefore facilitating further analysis in this area.

The second approach for dealing with a GP who will not release documents outside the fund's office is for the LP to review the documents at the manager's office during the on-site visit stage of the ODD process.

In certain cases, an investor may seek to have a private equity fund under consideration for investment complete a list of specific questions on a form commonly referred to as a due diligence questionnaire (DDQ). Many GPs will typically have some sort of DDQ already prepared. These GP-prepared DDQs will typically provide answers to commonly asked questions about the firm, such as historical asset

EXHIBIT 10.3 Common Documents Collected during an Initial Private Equity Operational Due Diligence Review

Document Type	Example Documents	Notes
Fund-specific legal documentation	Fund offering memorandum	
	Subscription documents	
	Articles of association (if applicable)	
	Limited partnership agreement (if applicable)	
	Fund formation documents	
Firm legal and compliance documentation	Firm formation documents	Another document commonly asked for associated with formation is a certificate of good standing, or equivalent, which in part shows that no outstanding taxes or liens are in place.
	Compliance manual	If not included in compliance manual: • Employee personal trading procedures • Electronic communication policy • Anti-money-laundering policies and procedures
	Code of ethics	
Financial documents	Audited financial statements	If no audits are available, previous fund audits are typically collected.
Marketing communications	LP investor letters	
	Samples of recent marketing materials	Examples include presentations (commonly called pitch books) and GP letters to LPs.
Other documentation	Firm organizational chart	
	Business continuity and disaster recovery plan	
	Valuation policy and procedures	
	Details of insurance coverage	LPs may also ask for actual copies of insurance certificates in addition to a summary of coverage.
	Operations policy manual	
	Information technology policies and procedures	

figures. In addition, some GPs may prepare DDQs addressing specific areas of fund operations, as well as more general DDQs focused on providing broader firm detail.

10.4 ANALYZING PRIVATE EQUITY LEGAL DOCUMENTATION DURING OPERATIONAL DUE DILIGENCE

During the document analysis stage, the documentation that was previously collected by LPs is now analyzed. As part of this process, the analysis techniques employed serve to meet several goals. Of course, the specific techniques used in analysis vary greatly among the document types being reviewed. As noted previously, the list of documents collected crosses several disciplines and areas of the firm.

For example, when analyzing financial statements, different skills and analytical techniques are employed as compared to when reviewing an administrator service provider agreement. While it is not practical to anticipate every document that may be collected and reviewed during the ODD process, there are common document type categorizations. In this chapter, we focus on the three most common document types that are typically collected and reviewed during the private equity ODD process:

1. Legal documentation
2. Financial documentation
3. Information technology documentation

We begin our analysis with legal documentation.

10.4.1 Two Primary Motivations for Designing Private Equity Legal Structures

Integral to a discussion of private equity legal documentation is to understand the private equity fund's legal structure. The two primary reasons for the design of private equity legal structures are:

1. To facilitate the implementation of tax efficiency
2. To limit liability among the entities involved

10.4.2 Common Private Equity Legal Structures

It is common for private equity firms to be organized under a combination of partnership entities. Leading this group of entities is typically the GP. In the context of legal documentation, the GP may commonly be referred to as the managing partner of a private equity company. A private equity fund may also have an intermediary-level entity known as the manager or investment adviser (IA). The IA usually sits between the general partner and investors, and may technically serve as the manager of a particular private equity fund.

Organized under the GP (or IA, as the case may be) is typically the private equity fund itself. One common private equity fund structure is a legal structure known as a limited partnership. Sitting below the private equity fund entity itself are commonly

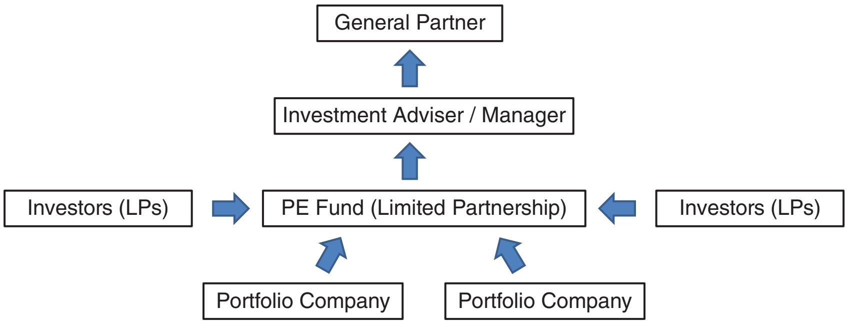


EXHIBIT 10.4 Example Private Equity Fund Legal Structure

the portfolio companies that the private equity fund invests in. LPs are the investors that feed into the private equity fund. An example of a private equity fund structure is outlined in Exhibit 10.4.

10.4.3 Understanding Offering Memoranda Functions

Across the spectrum of legal documentation, one of the key documents that is collected and reviewed from private equity funds is the offering memorandum (OM), commonly referred to as the private-placement memorandum (PPM). (These two terms will be used interchangeably in this discussion.) An OM is the central controlling legal document for the PE fund.

Due to its importance, we will focus our discussion of PE legal documentation analysis around the OM. Before discussing the common approaches employed in analyzing an OM, it is useful to provide some background on an OM document itself. This document is the central legal bible of the private equity fund under review. An **offering memorandum** or **private-placement memorandum** seeks to accomplish four key functions:

1. Limited partner education
2. Risk disclosure
3. Risk assignment
4. Assignment of decision-making authority

When analyzing an OM or a PPM, an investor may notice that the actual document itself is quite long, often exceeding 100 pages. One of the reasons for the length of these documents is that they provide a great amount of education to limited partners about a variety of issues. The OM includes factual information about not only the specific investment strategy of the private equity fund under consideration but also a number of related items, including key fund and GP personnel, as well as information about the GP organization itself. For private equity funds in particular, this factual background information may also cover information about funds from prior vintage years managed using the same investment strategy.

The second primary function of the PPM is to serve as a vehicle for risk disclosure. Many investors understand the risks common to investing in private equity, including the illiquidity concerns of long-term investments made by these funds.[2] The PPM serves not only to highlight these risks but also to provide a number of detailed disclosures related to such risks. In some cases, these disclosures may be broad enough to encompass a number of risks as well as to satisfy the technical goals of these disclosures. The technicality of such disclosures is often related to the risk assignment function of such disclosures.

Risk assignment in the context of fund documentation refers to anticipated ways for placing responsibility for different risks with different parties. Two common related legal terms contained in OMs that assign risk are *exculpation* and *indemnification*. **Exculpation** is a contractual term that relates to freeing someone from blame; **indemnification** relates to a duty to make good on a loss.

Exculpation and indemnification clauses are used in private equity PPMs to outline a limited set of legal standards at which predefined parties would be liable for certain actions or losses. Common legal standards employed include gross negligence, fraud, willful malfeasance, bad faith, and dishonesty. Following is an example of a typical indemnification clause—this one for the GP—that would be contained in a private equity fund OM:

> *The General Partner is liable to creditors for the debts of the Partnership. However, none of the affiliated parties, nor any person designated to wind up the affairs of the Partnership pursuant to the partnership agreement, will be liable for any loss or cost arising out of, or in connection with, any activity undertaken (or omitted to be undertaken) in connection with the Partnership, including any such loss sustained by reason of any investment or the sale or retention of any security or other asset of the Partnership, except for any liability caused by his, her, or its fraud, gross negligence, or willful misconduct.*

This indemnification clause outlines that the GP and affiliated parties would be liable for the debts of the partnership under a specific set of legal standards (i.e., fraud, gross negligence, or willful misconduct). In other circumstances, such as an honest error, the GP would not be liable to LPs.

Decision-making authority is perhaps best understood by way of example. Let's say a private equity fund makes an inadvertent fund accounting error, which causes the fund to be late in paying the legal bill of fund counsel. Let us further assume that as per its arrangement with fund counsel, all legal bills not paid in a timely manner accrue interest at a rate of 1.5% per month. The error in our example went unnoticed for three months, and therefore the fund had an obligation to the law firm that was 4.5% greater than it would have been had the error not occurred. Who should be responsible for paying the additional 4.5%?

On the one hand, the argument could be made that GP accounting personnel did not make the error intentionally and that it came up in the legitimate course of business. Such errors, it could be argued, do occasionally happen, and it should not be considered the GP's fault. This line of thinking would support the notion that the fund, and by association its LPs, should be responsible for this additional 4.5%.

On the other hand, it could be argued that it was the GP's fault that this error occurred and that the GP has a responsibility to the fund's investors to institute strong fund accounting controls to prevent such errors. If it had, the error would not likely have occurred. Therefore, the additional 4.5% expense should be absorbed by the GP at the management company level and not borne by the fund and its LPs. Which line of thinking is correct?

Suppose instead we believe that although such an error can occur in the standard course of business, the fact that it was not detected for three months is egregious. In this view, perhaps a compromise should be in place in which the fund is charged the additional expense of one month (1.5%) and the GP absorbs the remaining 3%. Would this be more equitable? Who gets to make such a decision: the GP or the investors?

As you can see, there are a number of considerations that may arise in practice when dealing with such issues. Carefully drafted PPMs can help by providing guidelines and mechanisms for resolving such issues before they arise. For example, perhaps the PPM makes clear that such decisions are the exclusive realm of the GP. Some PPMs may even address specific contemplated issues in advance—for example, clarifying that LPs would be responsible for trade errors made in good faith, or stating that there are caps in place on the total amount of fund expenses.

Whether you as an investor think the guidelines and mechanisms are fair is a different matter. The point is that the OM outlines these rules prior to an investor committing capital to the private equity fund. If an investor doesn't like the rules, the time to seek to negotiate certain risk assignment terms is generally prior to investing.

Investors can often request to enter into what is known as a **side letter** to negotiate such terms, which is an agreement between an investor and the fund that amends the OM to afford a specific investor with certain negotiated provisions. Examples of common items negotiated in side letters could be additional fund transparency or beneficial redemption rights. It should be noted that side letters do not apply to all LPs in a pooled fund structure—just those designated in the side letter.

Performing due diligence on these types of issues and negotiating any side letter terms prior to investing is critical, because there is often little room for negotiation after the capital has been committed.

10.4.4 Distinguishing between Legal Counsel and LP ODD Legal Document Reviews

Analysis of the actual terms of a PPM is a detailed process that requires knowledge of common document terms of private equity funds in general, as well as understanding the specifics associated with the particular fund under review. It should be noted that in the context of an ODD review, although the PPM deals with matters related to law, a review of the document by an LP may differ from a review by legal counsel. The goals of such reviews may be different. A review by an LP's law firm, for example, may be focused on detecting terms in the document that may violate certain investor guidelines or laws, such as the Employee Retirement Income Security Act (ERISA). An LP ODD review, however, may focus on risk analysis and term negotiation. An example of such a term is something known as a key person provision.

A key person provision typically provides LPs with options to redeem capital and not have any additional capital called should key individual(s) associated with the fund cease to be associated with the fund and GP. Because it is not illegal in and of itself for a GP's PPM to omit a key person provision, the absence of such a clause might not be cause for concern during review of the document by LP legal counsel. However, in the context of an LP ODD review, the omission of such a clause may raise concerns due to the risk considerations involved around key personnel of the firm. This is an example of the differing scope of such reviews and the benefits to LPs of incorporating third-party legal counsel reviews with LP ODD PPM reviews.

Returning to the key person provision, the following are five examples of common questions an LP would consider when reviewing this fund term in the context of an ODD review:

1. Is such a term present or omitted?
2. How many individuals(s) are deemed to be key persons?
3. What triggers key person conditions: death of a key person, incapacitation, or lack of involvement with the fund/firm for a predefined period of time?
4. Does the clause require multiple key persons to undergo a key person condition in order to trigger an event, or just one?
5. What is the notice period to investors once a key person event is triggered?

Integrating the risk implications of key fund legal terms such as key person provisions into the broader operational risk assessment process is where ODD can add value over a pure legal analysis of the legal terms of documents such as PPMs.

10.4.5 Analyzing Other Common Private Equity Offering Memorandum Terms

As noted previously, private equity PPMs are lengthy documents covering a wide variety of fund terms. Other key areas that are commonly focused on by LPs during a review of the OM, and associated considerations, are outlined in Exhibit 10.5.

Fund fees are also commonly analyzed as part of the PPM review process. In addition to a standard management fee, private equity funds may charge a number of other fees, including sales and distribution charges, fund start-up expenses, and other ongoing expenses. Another common private equity fee disclosed in the fund documents relates to what are known as management and performance fees.

Management fees are flat fees earned by the fund, regardless of the fund's performance, typically charged as a percentage of committed capital. During a review of a fund's OM, LPs review fee disclosures to confirm that the fees expressed by a GP in other materials, such as marketing presentations, conform with their expectations. GPs may charge a variety of other fees, such as deal fees, advisory fees, and monitoring fees paid by portfolio companies. Investors need to determine if those fees are earned by investors or by the GP, and whether those fees offset the stated management fees. Private equity funds also typically earn carried interest, which is sometimes referred to as performance fees or incentive fees. Carried-interest payment distribution schedules vary among PE funds. The amount and timing of carried-interest payments can provide valuable insights into other operational risk areas, such as employee compensation and retention approaches. When evaluating the language in

EXHIBIT 10.5 Key Areas of Offering Memorandum Considered during the Private Equity Operational Due Diligence Review

Area Reviewed	Comments
Fund and associated entity domicile	Common domiciles are Delaware for U.S.-based funds and the Cayman Islands for non-U.S. (i.e., offshore) funds.
Distributions	Distribution terms outline the ways, amounts, and timing of capital that is returned to investors. Commonly, 100% of distributions are paid directly to LPs until certain criteria are met, including recouping their original investment and a predefined rate of return. Distributions then typically proceed under a predefined schedule, with the GP sharing in a portion of the profits.
Capital commitments and unfunded commitments	These terms describe the process by which LPs commit capital to a fund, including minimum commitment amounts. OMs also typically contain terms that if capital is called from other LPs and they cannot meet the commitments, other LPs may be responsible for meeting these unfunded commitments.
Tax considerations	While tax considerations vary across jurisdictions, OMs will typically detail key tax considerations for the fund and LPs. Although the goal of this LP analysis is not to produce a formal tax opinion, gaining an understanding of the tax structure and anticipated consequences of a fund's investments should be considered during an ODD review.
Conflicts of interest	**Conflicts of interest** refer to the ability of the GP, funds, or key personnel to participate in other activities that may present a potential conflict with the fund under review. Three examples of common conflicts of interest that may be in place: 1. Allocation of investment opportunities among multiple funds 2. Conflicts relating to asset dispositions 3. Participation in outside business activities

the OM with regard to carried interest in the context of an ODD, review questions to consider include these four:

1. Are waterfall distributions or other similar schemes employed?
2. Is a deal-by-deal or fund-as-a-whole approach applied to carried interest?
3. Who determines which individuals or groups participate in carried-interest distribution?
4. What are the vesting periods for the payout of carried interest to team members?

In regard to fees in general, from an ODD perspective, LPs should be conscious of the timing of the collection of fees. Fees may be collected in one of two ways:

1. Fee collection in advance: This refers to fees being collected at the beginning of the period in which they are due (e.g., the beginning of the month).
2. Fee collection in arrears: This refers to fees being collected at the end of the period in which they are due (e.g., the end of the month).

Due in part to considerations related to the time value of money, it is generally considered advantageous to the LP for a fund to collect fees in arrears as opposed to in advance.

10.5 OPERATIONAL DUE DILIGENCE BEYOND LEGAL DOCUMENT ANALYSIS

Likely subsequent to analysis of legal documents, there are several additional stages of analysis that LPs should perform on a prospective PE fund investment.

10.5.1 Analyzing Private Equity Valuations in the Context of Legal Documentation

Valuation is an example of a risk area that relates heavily to the investment and operational areas of private equity funds. From an investment perspective, when fund managers seek to make an investment, understanding current and future valuations of the underlying investment targets certainly plays into their analysis. In an operational context, the analysis of valuations is focused on two key areas:

1. Understanding the GP's internal valuation policies and procedures
2. Analyzing the level of independent oversight in the valuation process

In the context of a legal review, an OM will contain a reference to the ways in which a fund will value its securities and will calculate a fund net asset value (NAV). The level of specificity in the OM will vary depending on the fund. Additionally, an OM is often written to provide the fund with a great deal of flexibility in regard to its approach to valuations. This can sometimes present a challenge to LPs seeking to determine how valuations are done in practice to value underlying PE holdings.

In order to complement the sometimes overly vague and legally technical language of PPMs, some GPs may maintain documents detailing separate valuation policies and procedures. Other documents, such as audited financial statements and even marketing materials, may provide further clarification regarding the actual valuation practices employed.

One common governance mechanism often used to facilitate internal GP oversight of valuation is a **valuation committee**, which typically consists of representatives of various departments, both investment and operational, throughout the firm. The valuation committee typically meets on a predefined recurring basis (e.g., quarterly) to review fund valuations and any proposed valuation changes. Two common reasons for adjusting the valuations of positions held are as follows:

1. Occurrence of a material event related to the security, most frequently write-downs for impaired firms or increases in valuations from later rounds of funding
2. Changes in overall markets that facilitate a revaluation of the security

As part of its review of portfolio holdings, a valuation committee may also ask to review internal pricing work and any associated pricing memoranda that may have been prepared. The following are three key considerations for LPs in reviewing the level of independence in the work of the valuation committee:

1. Does the committee membership appropriately represent the different parts of the firm, or is the committee heavily dominated by investment personnel?
2. What is the process by which the committee may challenge or revise valuations?
3. What level of independent analysis does the committee perform?

In certain cases, private equity firms may use third-party valuation agents to assist in producing valuations. Some PPMs will outline specific guidelines for the use of these third-party valuation agents. In some cases, their use may be completely at the discretion of the GP. In other cases, the fund may set minimum position size limits (e.g., 10% of fund NAV) before such agents can be used. It is often up to valuation committees to ensure that the rules for the use of such valuation agents are followed. It should also be noted that these valuation agents are different from third-party administrators, which are discussed in more detail later in the chapter.

10.5.2 Analyzing Private Equity Advisory Committees

A common feature of many private equity funds is that they contain a fund advisory committee, sometimes referred to as an advisory board, which typically consists of several large representative investors of a fund. These committees typically serve to provide notifications to committee members with regard to material operational developments in the life of a fund. Advisory committees may also be consulted by GPs with regard to the undertaking of certain investment opportunities or other material portfolio changes.

There are no legal requirements for funds to maintain these committees. If such a committee is present, the OM will typically describe the rights and duties of this committee. In some cases, an OM may outline that the role of the advisory committee goes beyond merely a notification function, instead outlining that the committee will be consulted with regard to certain regular items and new fund developments. Examples of the items that an advisory committee may be consulted on include these five:

1. New acquisitions and financing of fund investments
2. Dispositions of fund investments
3. Review of financial statements, appraisals, and valuations
4. Review of potential conflicts of interest
5. Appointment of a new auditor or other key service providers

It is important to note that although advisory committees may be consulted regarding such items, they may have no actual veto or approval power to prevent or approve certain proposed changes. The focus of these committees is typically to give the largest investors in a fund transparency and to promote a dialogue with GPs, as opposed to giving them power to force actual changes. As previously noted, the creation of an advisory committee is typically at the discretion of the GP. That being said, many investors prefer the enhanced transparency associated with these committees; therefore, they have become increasingly popular in recent years.

10.5.3 Private Equity Audited Financial Statement Review

The primary financial documentation reviews by LPs during the private equity ODD process are the audited financial statements of a fund, sometimes referred to simply as audits. These audits are prepared by a third party known as an auditor. In some cases, an existing fund is being analyzed for which audited financial statements are available. As detailed later in this section, in cases of new funds, financial statements may be available on funds of earlier vintage.

Depending on the jurisdiction in which the fund is based, the accounting standards used by the auditors may differ. Common accounting standards include Generally Accepted Accounting Principles (GAAP) and International Financial Reporting Standards (IFRS). Different countries that use GAAP may have their own specific versions, such U.S. GAAP or UK GAAP. The choice of accounting and audit standards not only influences the presentation of the audited financial statements but also governs the underlying assumptions used in preparing these statements.

Depending on the format used, the rules for the presentation of items—such as fund expenses—may differ. Additionally, the sections included in the financial statements may vary based on format. The common sections of audited financial statements that should be reviewed by LPs during the ODD process are outlined in Exhibit 10.6.

It should also be mentioned that the financial statements typically contain a notes section at the end that provides a number of important disclosures and additional

EXHIBIT 10.6 Common Sections of Audited Financial Statements

Audited Financial Statement Section	Comments
Opinion Letter	This section provides a summary of the auditor opinion of the financial statements. An auditor opinion that is qualified means that the auditor is certifying the statements subject to a qualification. In limited circumstances, **qualified audit opinions** may be used to designate acceptable deviations from accounting standards. In other cases, qualified opinions may be used to raise red flags to investors, which should be thoroughly vetted during the operational due diligence process.
Statement of Assets and Liabilities	Commonly known as the **balance sheet**, this section provides a summary of assets, liabilities, and partners' capital.
Statement of Operations	Commonly known as the **income statement**, this section provides a summary of income and expenses.
Statement of Cash Flows	This section outlines the movements of cash throughout the fund as well as financing-related cash flow activities.
Statement of Changes	This section, sometimes known as the statement of changes in partners' capital, outlines items related to partner capital allocations, contributions, and withdrawals.
Schedule of Investments	This section typically details portfolio holdings and may classify them in summary form according to predefined categories such as by sector or region.

information related to fund operations. The seven main areas typically covered by financial statement notes are:

1. Organization and business
2. Summary of significant accounting policies
3. Investments
4. Commitments and contingencies
5. Related-party transactions
6. Financial highlights
7. Subsequent events

A common consideration for LPs when investing in private equity is that an investor may be committing capital to a brand-new fund with no history or one that is the next series in a vintage-fund framework. This new fund typically has no operating history and therefore no audited financial statements. LPs can often overcome these concerns by reviewing the audits for funds of the previous vintage. Although not an exact indication of the financial state of the fund under review to which the LP is considering allocating capital, this review of similar funds can often provide useful insights into the overall management of finances of previous funds. Additionally, investors can gain familiarity with the format and style of the statements, which will typically use the same auditor across vintages unless a firm-wide or strategy-wide switch of the auditor takes place.

10.5.4 Information Technology and Business Continuity Planning/Disaster Recovery Documentation Review

Private equity funds, as with most modern fund managers, are heavily reliant on information technology (IT). There are several documents associated with IT that can be analyzed. Information technology documentation typically includes a review of various aspects of a fund's IT function, including policies related to hardware, software, and ongoing support. Private equity firms typically maintain documentation of IT policies and procedures that address the ways in which firms manage these issues. When reviewing these documents, common considerations for LPs include these five questions:

1. What is the organization of the firm's IT function?
2. What process is in place with regard to the rolling out of new revisions of IT software and hardware?
3. How does the firm approach information security issues, including protecting data, managing firewalls, and updating credentials for employees and consultants?
4. What escalation procedures are in place to manage IT issues that may arise during the course of the firm's operations?
5. What is the firm's IT hardware and software support management plan? If the PE firm maintains multiple offices, how is this process managed for the different locations?

An area heavily related to IT is business continuity planning and disaster recovery (BCP/DR), which refers to how a firm seeks to continue operations in the face of a disaster event (e.g., a natural disaster or terrorism) or a more temporary business disruption (e.g., the firm's office becomes temporarily inaccessible or inoperable). Although BCP/DR is not exclusively the purview of information technology, reviewing BCP/DR documentation with an IT focus in mind can certainly add value during the ODD process.

When reviewing BCP/DR documentation, a key consideration relates to gauging an understanding of the implementation and testing of technology policies and procedures. Testing refers to a firm conducting dry runs of simulated business disruptions or disaster events. PE firms' BCP/DR will typically describe policies for testing of the plans. The following are five examples of key considerations during a review of BCP/DR documentation:

1. Is testing described from a personnel perspective, technology perspective, or both?
2. How is employee contact information shared and updated throughout the firm so that employees may remain in touch in the event of a disaster?
3. How frequently is it anticipated that testing will be performed?
4. If any issues arise during the testing process, which GP groups will review the testing results?
5. How frequently will the firm revise its BCP/DR?

For private equity firms in particular, some investors question the benefit of analyzing BCP/DR documentation. This questioning is based on the fact that private equity funds do not typically trade as frequently as do hedge funds. The thinking goes that any business disruptions would not be as detrimental to PE funds due to this more limited trading activity. However, private equity funds do engage in a number of time-sensitive projects, which a business disruption could materially affect. For example, from an operational perspective, if a business disruption occurred during the quarter-end close of the accounting books of the fund, a delay could significantly influence the fund's ability to distribute investor statements. For these reasons, it is therefore considered best practice to incorporate a review of BCP/DR documentation into an ODD assessment of a private equity fund.

10.6 ON-SITE MANAGER VISITS

On-site manager visits form a crucial step in the operational due diligence process.

10.6.1 Selection of Visit Location

After the documentation review process is complete, the next stage in the process is for investors to go on-site with a private equity manager at the fund's office to conduct an in-person meeting. One consideration that may arise in setting the location for the office review is the question of what to do should the GP maintain multiple offices. In the context of setting a meeting location focused on operational risk, it is commonly considered best practice to focus the on-site meeting where the majority of fund operational procedures are performed. Of course, it is not unheard-of to visit multiple locations should it be deemed necessary to meet face-to-face with individuals in multiple offices.

10.6.2 Desk Reviews Are Not Best Practice

Some investors dangerously seek to cut corners and omit the on-site visit during ODD. This type of ODD review is a **desk review**, which is limited to a review based solely on documents collected and perhaps conference or video calls. The desk review approach is not considered to be best practice. Three reasons LPs may argue in favor of desk reviews are the following:

1. Lower cost of desk reviews compared to on-site visits
2. Shortened overall review time for desk reviews
3. Belief in equal information collection by both review processes

Although desk reviews may result in lower costs and shorter ODD review processes, not including an on-site visit with the GP during PE ODD typically results in a less comprehensive review. This can subsequently expose investors to increased levels of operational risk.

Earlier in our discussion, we referenced certain documentation that GPs may decide not to allow LPs to review outside their offices. We outlined two techniques by which LPs typically overcome this hurdle. The first is to collect compromise documentation. This option is still feasible under desk reviews. The second option is to review documentation on-site with the manager. Because no on-site visit is conducted under a desk review, this option would not be available to LPs following a desk review approach. This represents another deficiency of desk reviews as compared to full ODD reviews.

Additionally, due to the enhanced illiquidity of private equity investing, investors who do not conduct on-site visits may later discover operational risks from which they have limited mechanisms to protect themselves and may learn that they are unable to facilitate improvements at the fund after making an initial capital commitment. Consequently, it is considered best practice to conduct an on-site visit as part of the ODD process to facilitate a more up-front, comprehensive review, as the long lockup of PE funds prevents redemption for most problems that may arise after the capital has been committed.

10.6.3 On-Site Visit Agenda Development

Before conducting the actual on-site visit with a manager, one of the first steps typically taken by investors is to develop an agenda for the meeting. It is considered best practice for LPs to be proactive in developing an agenda as opposed to letting the GP dictate the schedule. The information reviewed by LPs during the document analysis stage is often useful in this regard. Expanding on this point, four key goals for LPs to consider in developing an on-site agenda are:

1. Verifying information contained in the previously collected documentation
2. Filling in gaps in documentation
3. Going beyond documentation to add further details regarding operational practices and procedures
4. Analyzing the firm's ability to demonstrate operational practices

In addition to these goals, there are two primary overarching aims of the on-site agenda. The first is to establish the topics that an LP wishes to cover during the meeting. In practice, many of the items covered during the on-site visit will parallel the topics addressed during the document collection process. The following list presents seven common items LPs may want to cover in developing an agenda for an on-site visit:

1. General firm and fund overview, covering historical assets under management, current and previous funds' investor bases, personnel, and fund structures
2. Fund accounting and cash movement and management procedures
3. Valuation process and procedures
4. Overview of the legal and compliance function
5. Discussion of the ways in which the fund and firm work with service providers
6. Information technology policies and procedures
7. Business continuity planning and disaster recovery

This is only a representative list, and the specifics of each on-site agenda should be customized as applicable to the GP under review. For example, a private equity fund may completely outsource the fund accounting procedures to a third party. In these cases, the agenda would be tailored to discuss the ways the PE firm manages this third-party relationship as opposed to focusing on the actual nuts and bolts of fund accounting procedures, details that could be reviewed separately through discussions with the outsourced service provider. Additionally, an LP may wish to cover topics not appearing on the list, such as operational considerations in the firm's approach to risk management, which could certainly be added to the agenda as well.

The second primary goal of the on-site visit agenda is to meet face-to-face with key operational personnel, who typically include the following seven roles:

1. Chief financial officer (CFO)
2. Chief compliance officer (CCO)
3. Chief operating officer (COO)
4. Chief technology officer (CTO)
5. Fund controllers and accountants
6. Deputy compliance officers
7. Investor relations personnel

It should be noted that not every GP will necessarily maintain individuals with these titles. In practice, certain individuals may also wear multiple hats, such as a shared CFO/COO. The goal in developing this list of personnel to meet with is not to focus so much on each individual's specific title but to ensure that the appropriate information will be covered with each person.

10.7 EVALUATING META RISK

The April 15, 2008, report of the Investors' Committee of the President's Working Group on Financial Markets titled *Principles and Best Practices for Hedge Fund*

Investors references the concept of **meta risks**, which are defined as "the qualitative risks beyond explicit measurable financial risks. They include human and organizational behavior, moral hazard, excessive reliance on and misuse of quantitative tools, complexity and lack of understanding of market interactions, and the very nature of capital markets in which extreme events happen with far greater regularity than standard models suggest."

In an operational risk context, we can think of meta risks as the catchall category used to account for all non-investment-related risks not covered by a particular category. Examples of meta risks could be the risks associated with a fund manager's expenditures of management company funds on expensive office decorations rather than on hiring additional staff. While it is certainly a manager's right to make such expenditures, some investors may argue that such exorbitant expenditures represent operational risks that should be considered during ODD. Another example would be a fund manager who is confrontational or defensive during an on-site meeting. Once again, some investors may feel that such a poor attitude should be considered as part of the overall evaluation of the fund manager. As these types of risks do not necessarily fit neatly into any one specific operational risk category, the meta risk area is used to categorize them.

The process of assigning values to meta risks is inherently subjective and can vary from investor to investor. However, it is an area in which a highly skilled operational risk professional with a deep understanding of operational trends in the marketplace can add direct value.

10.8 FUND SERVICE PROVIDER REVIEW AND CONFIRMATION

The next stage in the ODD process is to interact with the fund's service providers. Service providers are third-party firms that provide a variety of services, both investment-related and non-investment-related, to the firm and funds. Investors have three primary goals when interacting with service providers:

1. Independently confirm the GP's relationship with service providers.
2. Gauge appropriateness of the service providers for the GP; some common questions to consider in this regard include the following:
   - Are the service providers of appropriate size for the GP, or are they too big or too small?
   - Are the fees being charged appropriate for the level of services being provided?
   - Similarly, are the providers appropriate for the GP at their current stage of development (years in business, assets under management)?
3. Analyze the quality of the services being provided to the PE firm and fund.

Common fund and firm service providers are summarized in Exhibit 10.7.

Other common service providers may include custodians, insurance brokers, utility providers for items such as telephony or power generation, personnel recruiting firms, preemployment screening firms, and third-party service providers such as valuation agents, risk analysis services, and marketers and fund distributors.

EXHIBIT 10.7 Common Fund and Firm Service Providers

Service Provider Type	What They Do
Bank	Banks hold cash for both firm and funds. They also typically can provide standard banking services such as checking and processing incoming capital subscriptions.
Information technology consultant	IT consultants can provide a number of services to firms, including customized software projects, hardware management, and ongoing IT support.
Legal counsel	Third-party legal counsel (i.e., law firms) can assist the fund and management company with a number of law-related services. These can range from establishing and maintaining procedures to complying with local employment laws, and to more investment-related areas such as deal structuring and regulatory filings.
Compliance consultants	Compliance consultants typically work to augment in-house firm compliance functions. These groups can perform a number of compliance-related services, including mock regulatory audits and ongoing training on compliance-related matters.
Auditors	The primary goals of auditors are to prepare audited financial statements for the funds. Other services may include management company accounting and tax advice.

Additionally, as part of the service provider review process, LPs may reach out to other more investment-related service providers, such as trading counterparties. One example would be a swap counterparty that may be used by GPs. From an ODD perspective, the reviews of these counterparties will typically focus not on the investment merits of any interaction with such counterparties but on the operational interaction between the GP and the counterparties, as well as the terms of such arrangements.

Private equity firms may also work with a third-party service provider known as a fund administrator. Fund administrators provide fund accounting and shareholder services. Fund accounting typically includes independently reviewing portfolio pricing, reconciling cash and positions of the fund, and maintaining the official books and records of the fund. Shareholder services focus on overseeing the investor's capital commitment and withdrawal process and preparing investor statements. Due to the nature of additional oversight, particularly with regard to valuations provided by administrators, it is considered in the best interest of investors for third-party administrators to be used. However, unlike the hedge fund space, many GPs still administer their own funds under a process known as self-administration.

10.9 ONGOING PRIVATE EQUITY MONITORING CONSIDERATIONS

Ongoing operational due diligence monitoring refers to the process of conducting continued operational risk assessments of a fund after an initial review has already been completed and a commitment has been made. A comprehensive initial ODD review serves as a road map by which ongoing monitoring can be facilitated. Three common areas covered during the ongoing monitoring process are the following:

1. Reviewing the status of any planned operational improvements covered during the initial ODD review
2. Analyzing any new operational firm developments, such as the installation of new fund accounting systems and any changes to service providers
3. Evaluating the firm's current operational position since the time of the initial review in regard to continued operational scalability and controls

Operational scalability refers to the firm's ability to build on existing systems in order to continue to support growth in an organized manner, including via the addition of new resources, without material disruption. For example, operational scalability would raise concerns if sufficient fund account personnel were not added to support new fund launches.

Traditionally, there have been objections raised against performing ongoing ODD on PE funds. One of the primary reasons for this has been that once LPs had committed capital, barring certain unique circumstances, the LPs would generally be unable to redeem their investment or cease their obligation to fully fund the original capital commitment. Therefore, the question was raised as to what benefit ongoing monitoring provides if LPs were unable to act on it.

There is increasing appreciation of the need to perform ongoing ODD not only to facilitate LP risk management but also to enhance investor education and the monitoring of operational practices. Accordingly, there has been a trend toward increased ongoing ODD PE monitoring. Also contributing to this reversal are the more frequent ongoing dialogues between LPs and GPs in the periods beyond the initial capital commitment.

Additionally, the increased oversight by common groups of large LPs and mechanisms such as advisory boards have facilitated ongoing discussions with GPs regarding operational risks. Finally, rising liquidity in the secondary market for LP interests may allow investors who find material negative developments during their ongoing ODD process to use that information to their benefit by liquidating their holdings in a PE fund.

10.10 CONCLUSION

Of all asset classes, private equity funds display the most dispersion of the return. The differences between top and bottom managers are significantly greater than those of traditional managers, as well as other alternative asset classes. This is one reason that due diligence assumes such an important place when it comes to private equity investments. This chapter has discussed the process that should be followed by investors in performing operational due diligence on their private equity investments. Operational due diligence refers to the process of analyzing operational risks, which can be thought of in part as those risks that are not purely investment-related in nature and that arise from the daily management and business operations of the fund. These operational risks run the gamut from traditional back office trade operations to counterparty- and compliance-related risks.

Addressing the unique aspects and challenges associated with performing an operational due diligence review of private equity, this chapter has discussed the development of a flexible comprehensive operational due diligence program for private equity. It has included techniques for analyzing fund legal documents and financial statements, as well as methods for evaluating operational risks concerning valuation methodologies, pricing documentation, and illiquidity concerns.

NOTES

1. For further information on due diligence process for private equity funds see Scharfman (2012).
2. See Ang (2011).

REFERENCES AND FURTHER READINGS

Ang, A. 2011. "Illiquid Assets." *CFA Institute Conference Proceedings Quarterly* 28, no. 4 (December): 12–22.

Phalippou, L. 2011. "An Evaluation of the Potential for GPFG to Achieve Above Average Returns from Investments in Private Equity and Recommendations Regarding Benchmarking." Report to the Norwegian Ministry of Finance (February).

Scharfman, J. 2012. *Private Equity Operational Due Diligence.* Wiley Finance Series. Hoboken, NJ: John Wiley & Sons.

———. 2014. *Hedge Fund Governance: Evaluating Oversight, Independence, and Conflicts.* New York: Academic Press.

Shanahan, J., J. Marshall, and A. Shtekhman. 2010. "Evaluating Private Equity." Vanguard Group.

Steers, H. 2010. "Private Equity Fund Manager Due Diligence and Selection." In *The Definitive Guide to Private Equity Investment Fund Due Diligence*, 2nd ed. London: PEI Media.

CHAPTER 11

Private Equity Investment Process and Portfolio Management

There are challenges at all steps of the private equity (PE) investment process, and no process is ideal or optimal in all cases.[1] The most appropriate investment process for a given manager depends on that manager's objectives and tolerance for risk. This means that trade-offs are inevitable. The idiosyncrasies of private equity (e.g., its long-term nature, its illiquidity, and the peculiarity of its risks) mean that a unique balance must be found between the various components of the process, as illustrated in Exhibit 11.1.

Modern portfolio theory (**MPT**) is based on Nobel Prize–winning economist Harry Markowitz's insight that because they have unique risk and return characteristics, less than perfectly correlated assets can be combined in a way that maximizes return for any given level of risk. MPT relies on the fundamental principle of diversification and suggests that allocation choices are simple mean-variance efficient portfolios in perfect markets. Investors choose the appropriate combination of the risk-free asset and the market portfolio to create optimal portfolios that maximize expected return in line with their level of risk aversion (i.e., mean-variance efficient portfolios).

Despite its importance in finance in general, MPT's powerful mathematical apparatus is particularly difficult to apply in the context of private equity, where a sufficient quantity of data of a quality necessary for precise calculation simply does not exist. Risks, particularly those associated with investments in limited partnership funds, need to be modeled based on qualitative as well as available quantitative data. Realistically, portfolio management in private equity is about "satisficing"—a term coined by Herbert Simon, another U.S. Nobel laureate (1978)—which contrasts with optimum decision-making. Instead, we need to rely on heuristics for searching through available alternatives until an acceptable solution is found.

The scope for an institutional investor to conduct active management of long-term and illiquid PE assets is very limited; therefore, the investment process needs to rely on an ex ante design of the portfolio in order to achieve a desired risk and return profile. This risk and return profile describes the range of the PE portfolio's distribution of outcomes, while considering diversification effects and interaction of individual asset portions within an overall portfolio.

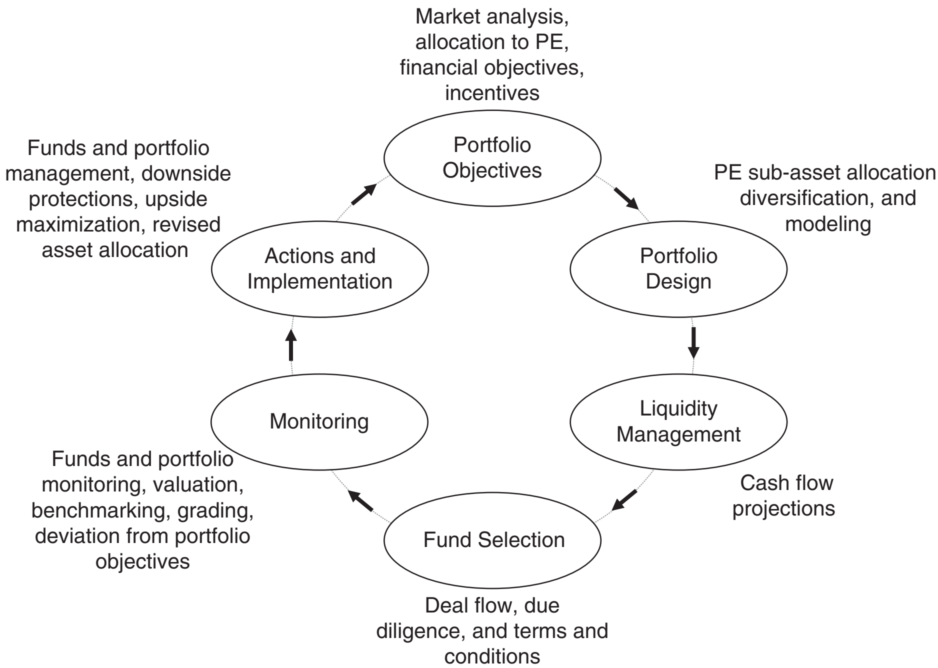


EXHIBIT 11.1 Investment Process

11.1 INVESTMENT PROCESS

The main decisions to address in the institutional investment process include strategic asset allocation, fund selection, level of diversification, and liquidity management. The challenge of managing an investment program requires an appropriate balance between an efficient fund selection that drives return and an effective allocation of capital and investment that drives exposure. Being highly selective by investing in only a few top-quality funds would maximize the expected returns stemming from these investments but would leave a major part of the resources dedicated to PE underinvested. A highly diversified allocation smooths the cash flows and allows for a nearly full investment of the capital allocated to private equity, thereby lowering the risk of receiving potentially extreme negative returns but at the expense of potential exposure to extreme positive returns.

For private equity, the three key performance drivers are fund manager selection and monitoring (see Chapter 9), the management of liquidity (see Chapter 13), and portfolio design in line with the targeted risk and return profile.

Because of the idiosyncrasies and lack of data related to private equity, it can be difficult to quantify the risks involved. Partially due to the opaque nature of the industry, not all outcomes are known, information is difficult to assemble, and the quantity and the quality of data are generally very poor. This is most notable in the case of technology-focused venture capital (VC) funds, which face rapid industry advances and change, new and evolving business models subject to constant forces of creative destruction, short boom-to-bust periods, and long investment periods, making it very difficult, if not impossible, to systematically collect data with any statistical merit.

That said, there are a significant number of assessable and even measurable factors in private equity (more than are typically perceived) that can be used in place of quantitative risk measures. A clearer differentiation between risk and uncertainty can be a useful tool for the management of a PE investment program.[2] Risk exists when a probability of occurrence can be reliably ascribed to an event, whereas uncertainty exists when there is no objective way to determine the probability. In typical discussions of modern portfolio theory, there is little differentiation between risk and uncertainty, as distributions of returns are assumed to be known (e.g., normal distribution). In private markets, however, investors are exposed to different degrees of uncertainty. Although a precise quantification is nearly impossible except in very specific situations, experts can associate experiences and opinions on various dimensions with categories for return expectations. They can also put lessons learned into a new or changed context and extrapolate from their experience. For these categories, an approximate quantification can be estimated and used to manage the uncertainty-return trade-off. For example, experts might grade funds subjectively with regard to risk and return, and then use the grades to select an appropriate aggregate risk level.

11.1.1 Portfolio Objectives

The starting point of any investment process is the definition of the objectives. These objectives should be set, or at least agreed on, by the investor (or authorizing entity). Identifying trends in the market and the amount of capital that can be profitably invested is crucial for establishing the investment strategy and setting targets. Construction of the portfolio requires the identification of attractive sectors and investment strategies. It is important that investment program managers receive the proper incentives to align their interests with those of the authorizing entity. The result of this analysis is a formalization of portfolio objectives that should ideally include the level of allocation to private equity, the targeted PE markets, financial objectives (e.g., desired risk-return profile) and nonfinancial objectives (e.g., environmental and social responsibility considerations), and how to achieve them within the context of a strategic asset allocation.

Asset allocation answers what asset classes to invest in and what would be an optimal mix among them. A large body of economic research suggests that asset allocation is the main driver of the investment performance of well-diversified portfolios and that the added value provided by the selection of a particular asset within the class is minimal. The fundamental justification supporting the relative importance of strategic asset allocation is the notion that individual asset classes perform differently in various market and economic conditions, and, as a consequence, this type of diversification is viewed as the only "free lunch" available in financial markets.

It is almost impossible to use standard risk-return optimization models to determine the appropriate allocation to private equity, because it is difficult to estimate the correct risk premium and correlation with other asset classes, in part due to the smoothed and infrequent valuations of the underlying portfolio companies. Also, an analysis of the correlations between private equity and other asset classes is not possible without making significant adjustments, such as computing PE returns under an assumption that intervening cash flows are invested in public market indices. In addition, private equity is lumpy, and desirable investments may not be available at the precise moment that an asset allocator decides to construct an optimal portfolio of private equity investments. Given the lack of information, most investors have to compromise and use naïve approaches, or a combination of quantitative optimization and naïve approaches, to construct portfolios.

11.1.2 How Much to Allocate to Private Equity?

How private equity is embedded in a broader asset allocation with, for example, publicly quoted assets is discussed in a number of publications, such as Swensen (2000), Fraser-Sampson (2006), and Cornelius (2011).

As the detailed discussion in Chapter 3 indicated, the Yale model has attracted wide attention among practitioners and the public alike. Also known as the "endowment model" of investing, it was developed by the Yale University endowment's David Swensen and Dean Takahashi. It is described in Swensen (2000) and is said to have embraced the principles of MPT, albeit in a simplistic and robust way. The model is based on diversification across asset classes with low correlations to maximize risk-adjusted investment return.

Some elements of the Yale model are consistent with a naïve diversification approach. The model divides the universe of available asset classes into a few categories and then proceeds to allocate to these asset classes using a combination of quantitative methods and naïve approaches. Since allocation to illiquid asset classes is an essential aspect of the endowment model, as previously discussed, naïve approaches to asset allocation may become a necessity. The portfolio should be rebalanced on a regular basis back to the original asset class weightings. Such naïve diversification recognizes that the expected risk-return relationship for long-term-oriented assets cannot be effectively quantified. By rebalancing, investors in theory would be selling private equity positions on secondary markets when prices have risen—or, to say it differently, when discounts are lower—and buying low (i.e., committing to funds when there is a scarcity of investor funding and portfolio companies can be acquired at attractive prices).

Central to the endowment model is an equity orientation to the detriment of asset classes with lower expected returns, such as investment-grade fixed income and commodities. The novelty of this model was that liquidity is to be avoided rather than to be sought out, since it comes at a heavy price through lower returns and has relatively high exposure to alternative asset classes—including private equity—compared to more traditional portfolios. That is, the allocation to illiquid assets should be maximized if these assets are expected to earn higher returns than liquid assets. Swensen (2000), however, gives no in-depth guidance on how the private equity portfolio itself is to be built.

While some institutions approach this by capping their PE exposures at around 5% to 10%, for some larger and more experienced foundations the allocation to private equity may exceed 30%. This naïve allocation can be estimated based on the analysis of four basic dimensions:

1. If the allocation is not large enough, the returns will not allow for the establishment of a dedicated team, which is a required condition if one expects to achieve above-average performance. Conversely, if the allocation is too large, it

may not be possible to find sufficient investment opportunities, or it may drive performance down if lower-quality funds need to be selected.

2. If the allocation is not large enough relative to the overall portfolio, it will have an insignificant impact on the overall portfolio performance (i.e., fail to "move the needle"). If the allocation is too large compared to the overall portfolio, the investor risks being underdiversified and overexposed to risks specific to private equity, such as illiquidity.
3. The benefit of adding private equity must be analyzed in the context of the existing portfolio. The goal is to determine the various risk dimensions that investors would like to diversify (e.g., industry sector, country, or company size) and to assess the impact of the PE allocation on these dimensions within the overall existing portfolio. For example, a high-tech early-stage allocation will provide a valuable diversification impact on a portfolio invested mostly in mature firms operating in the old economy.
4. The institution's liquidity needs should also be considered as a component of the analysis. This constraint essentially depends on the institution's regular liquidity needs to support its ongoing business, an important consideration for endowments and foundations. The lower the ongoing liquidity needs and the higher the excess capital, the more funds the investor can allocate to private equity, which is the case for the largest endowment investors. An investor with high regular liquidity needs and no excess capital is in no position to launch a PE program. In general, the more mature a PE program, the more liquidity it is likely to be generating, increasing the funds available for allocation to private equity.

Finally, it is useful to consider other institutions' allocations as a benchmark to supplement the information gathered by the portfolio manager regarding the attractiveness of private equity compared to other asset classes (Bain & Company 2015).

11.2 PRIVATE EQUITY PORTFOLIO: DESIGN

After having decided how much to allocate to private equity, the next question is to determine the appropriate allocations to the various PE market segments and to estimate the optimal level of diversification.[3] Some practitioners believe that PE portfolios should be managed more like public equity portfolios, drawing on many of the same tools and accepted principles while adjusting for market specificities. MPT suggests that underdiversified portfolios have a higher risk without adequate compensation in expected returns. However, as explained previously, it is questionable whether MPT can be successfully applied to real-world investing, specifically to portfolios of private equity assets.

MPT employs rigorous mathematical techniques for designing portfolios. For the model to work, one must be able to accurately quantify each asset's expected return and volatility of returns, as well as the correlations of the return of each asset relative to the return of all other assets in the portfolio. Adding to the challenge, while public market managers have more reliable statistical data to support their analysis, private equity and, in particular, VC managers lack such data. Indeed, the analysis of PE returns, volatility, and correlations is limited by the relatively short time series of

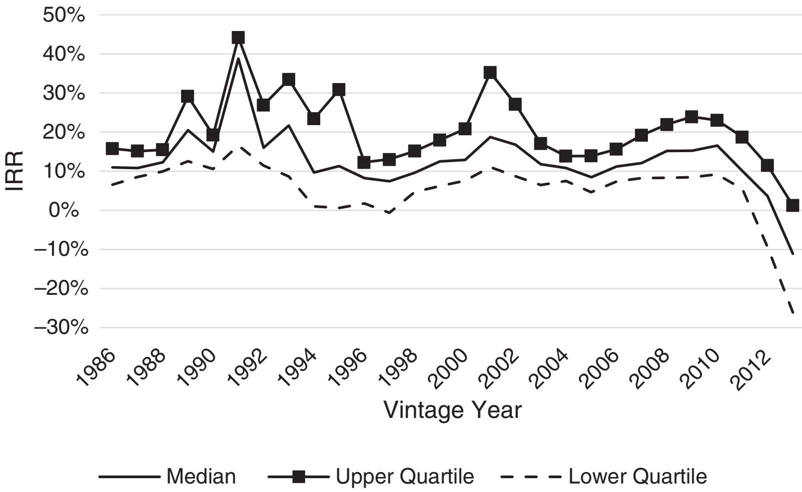


EXHIBIT 11.2 U.S. Private Equity and Venture Capital Index Performance as of December 2014
Source: Cambridge Associates, 2014c.

available performance data sets. Those data are also not fully representative of the market (i.e., incomplete) and are subject to various biases (see Chapter 8).

Most of all, data do not fully capture the uncertainty inherent in innovation-driven asset classes such as VC, since relevant past data would rarely be available on investments that emphasize new concepts. Moreover, MPT assumes a normal return distribution, which clearly does not hold for investments, especially private equity. In fact, the distribution of PE returns departs significantly from the normal distribution. Empirical results on private equity indicate large standard deviations of period returns, as well as significant skewness and excess kurtosis.[4]

11.2.1 Fund Selection

Gaining access to the top performers is believed to be critical to the fund selection process. Exhibit 11.2 and Exhibit 11.3 display relative performances of upper and lower quartiles of private equity funds by vintage year. As can be seen, for some vintage years, the differences between the internal rates of return (IRRs) of top-quartile and bottom-quartile managers exceed 15%. In fact, the average difference for U.S. funds is 15%, while the average difference for global ex U.S. funds is 20%. Therefore, the skill to select top-quartile funds is considered the core competence of PE specialists and is seen as a key performance driver for generating attractive returns for a portfolio of funds EVCA (2014) and Harris et al. (2015). The fund manager selection process is discussed in more detail in Chapter 9.

In the early days of the private equity industry, there were few players in a largely underexplored (i.e., inefficient) PE market that offered rich pickings. Access and due diligence were seen as everything, but now that the landscape has become increasingly competitive, these traditional tools often no longer lead to their desired outcomes. The institutional learning about PE investments and improvements of skill in

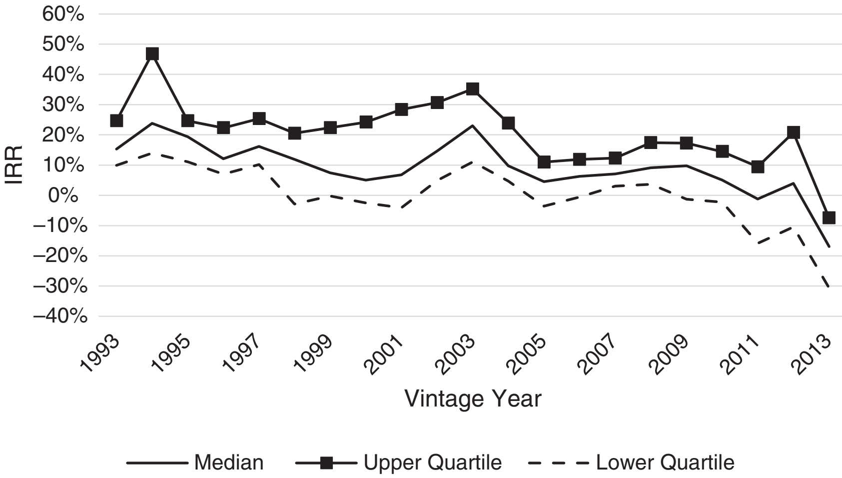


EXHIBIT 11.3 Global ex U.S. Private Equity and Venture Capital Index Performance as of December 2014
Source: Cambridge Associates, 2014, 2014b.

assessing profitable opportunities have, to a large degree, been neutralized by large increases of capital inflows seeking exposures to the asset class. While the investment capacity of private equity is likely to be very high in the long run, short-term increases in fund flows will inflate prices paid for available investment opportunities, resulting in low subsequent returns. In the midst of increased capital flows and compressed returns, investors may be tempted to think that their managers are above average and thus will not be affected by prevailing market conditions. Rather than relying on what are often incorrectly perceived as above-average selection skills, investors should focus on managing all aspects of the investment process well, including the evaluation of overall conditions of the industry.

11.2.2 Monitoring

Limited partners (LPs) should monitor the composition of their portfolios, identify trends within the PE markets, and evaluate the economic environment on an ongoing basis. For an LP, monitoring should include the specific fund investments as well as the overall composition of the relationships managed. Only through ongoing monitoring can an LP make informed, proactive portfolio decisions.

Analyzing concentration across all partnerships in the portfolio (e.g., by industry, investment style, stage, geography, vintage year, and cross-holdings between funds) can provide valuable insights. Stress tests can provide early-warning signals and increase the transparency of the portfolio. Tracking overall commitment level, contributions and distributions, return on investment to date, and expected final return on investment is also important for portfolio management. Identifying overexposures may lead to a review or an adjustment of investment objectives and, where severe imbalances exist, may require active management (e.g., secondary market transactions) to mitigate such overexposures.

At the individual fund level, compliance with contractual terms and investment style needs to be tracked. Fund monitoring is based on regular meetings with all parties involved and should include the following six activities:

1. Tracking planned versus implemented strategy
2. Reviewing the fund's financial investment, valuation, and divestment information
3. Analyzing the impact of relevant market trends
4. Assessing the risk of both individual investments and the overall portfolio
5. Measuring/benchmarking performance (see Chapter 8)
6. Verifying legal and tax compliance

Monitoring may also be relevant for liquidity planning (see Chapter 13). A more detailed description of the monitoring process is provided in Chapter 9.

11.2.3 Liquidity Management

Liquidity management is one of the key performance drivers in private equity. (See Chapter 13 for a description of the various approaches used.) It can be a difficult task to put money to work efficiently (i.e., carry minimal cash balances) while maintaining balance in the portfolio composition and preserving the quality of the individual fund investments.

Just looking at the capital invested in portfolio companies does not give the full picture of a fund's performance. Investors are rightly concerned with the total return on all resources dedicated to private equity, which should include undrawn capital (or at least the additional liquidities needed for the program). If a large part of the capital remains uninvested or parked in low-returning assets such as Treasury bills, the resulting drag on total return can be significant. To keep a program permanently and fully invested in portfolio companies, the so-called overcommitment strategy may need to be employed, whereby more commitments are signed than can be met with existing capital resources (see Chapter 13).

Funds of funds may certainly lose investors if they fail to deliver sufficiently attractive returns, but they can also go bust; in fact, some have already experienced significant problems in this area because they have not mastered the management of liquidity required to run such overcommitment strategies. There is anecdotal evidence that funds of funds have been struggling with this issue, as their overcommitment strategies have not worked out. The high degree of uncertainty regarding the timing of cash flows renders fund investing and liquidity management exceptionally challenging.

11.2.4 Actions and Implementation

The next step in the investment process is the implementation of portfolio management decisions. Active management is constrained because PE assets are long-term and illiquid, and offer only limited (in quality and over time) opportunities to increase investment or to sell. Primary investing, coinvesting, secondary investing, restructuring of funds, and securitization are possible means of actively managing the portfolio. The main purpose of fund restructuring is to stop value destruction rather than to create new opportunities. In practice, these tools can be applied only sparingly, as rebalancing the portfolio through buy-and-sell transactions can be very expensive and opportunities relatively scarce.

11.3 PRIVATE EQUITY PORTFOLIO: CONSTRUCTION

Even where the broad approach appears to be clarified, there remain a number of questions regarding the practical implementation. Private equity portfolios are constructed either from the bottom up or from the top down, or by using a combination of the two approaches.

11.3.1 Main Approaches

Approaches to constructing private equity portfolios are usually described as either bottom-up or top-down. The **bottom-up approach** is based on fund manager research, in which the emphasis is on screening all investment opportunities in the targeted PE markets and picking the perceived best fund managers. A **top-down approach** analyzes the macroeconomic conditions surrounding the targeted PE markets and then determines the weights and the combination of industry sectors, countries, fund styles, and so on that are best for meeting the PE program objectives under the likely scenarios.

While appearing to be in opposition, the bottom-up and top-down approaches are complementary and are typically used in tandem. This method, called the **mixed approach**, either starts with a bottom-up strategy, to which increasing top-down optimization is added, or starts as an iterative short process cycle, in which bottom-up screenings are followed by top-down analysis and then by bottom-up screenings.

Finally, investors who claim to have a top-down, bottom-up, or mixed approach all stress the importance of taking a proactive approach to deal sourcing. Investors cannot just wait for investment opportunities to arrive but must continuously search for the funds that fit their investment strategy and start making contact with these funds before they go to the market to raise capital. This requires a constant monitoring of the markets.

11.3.2 Bottom-Up Approach

Investors generally follow a bottom-up approach or a mixed approach, as it is widely believed that the quality of the fund management team is the most essential criterion, much more significant than such factors as sector or geographical diversification. The starting point of a bottom-up approach, also called the screening technique, is the identification of suitable investments, or those perceived to be the best (see Exhibit 11.4). This is followed by intensive analysis and due diligence in order to rank the funds by their attractiveness and identify which managers are the likely top performers. The investor conducts the due diligence necessary to select the manager, and then negotiates the structuring of the limited partnership agreements as well as the inclusion of covenants regarding the postcommitment monitoring, leading to the final fund investment decision.

The bottom-up approach has several compelling features. As it depends solely on ranking, this approach is simple, easy to understand, and robust. It enhances the expected performance by concentrating the portfolio in the highest-alpha funds (i.e.,

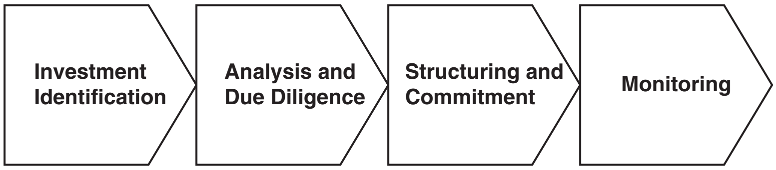


EXHIBIT 11.4 Bottom-Up Approach

funds with the highest expected performance uncorrelated with the market), while controlling for risk by diversifying across multiple funds.

However, the bottom-up approach is not without problems. As it is very opportunistic, it can lead to an unbalanced portfolio (e.g., a portfolio excessively concentrated in a specific sector or geography) carrying considerably more risk than intended, or can miss some important macroeconomic changes or opportunities (e.g., an aging population or renewable energy).

11.3.3 Top-Down Approach

The top-down approach takes the big picture as its starting point as opposed to individual fund selection (see Exhibit 11.5). Investors who follow a top-down approach place more emphasis on the management of the strategy, asset allocation, portfolio diversification, and macroeconomic considerations.

The top-down approach is a process that analyzes the macroeconomic conditions surrounding the targeted PE markets and attempts to determine the strategic asset allocation (i.e., the combination of industry sectors, geographies, and fund styles that are expected to be the most likely to meet the PE program objectives under the likely scenarios). The main criteria used in the evaluation process are political, economic, and currency risks. Other criteria examined include, for example, the extent to which each particular market has accepted private equity as a form of financing and investment, and the degree to which the environment is conducive to entrepreneurial activity. In this context, the investor takes into account various factors influencing the ability to invest, such as due diligence standards, accounting and tax issues, and the enforceability of legal rights (this last issue is particularly relevant in the case of many emerging markets). Finally, the availability of both attractive investment opportunities and exit opportunities for investments is considered. In the broadest sense, the decision to allocate commitments to vintage years could also be considered part of a top-down approach. After establishing the strategic asset allocation, the

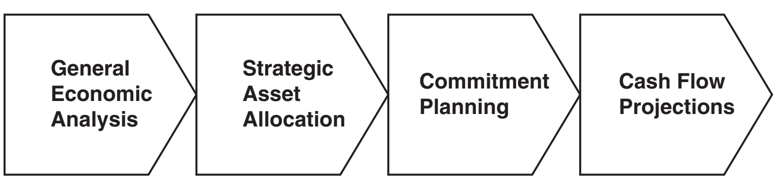


EXHIBIT 11.5 Top-Down Approach

investor determines what resources will need to be committed to the fund. This commitment planning depends on the investor's desired exposure level, risk tolerance, and available resources for investing. The final commitment strategy is determined based on cash flow projections and stress testing, as well as choosing funds that best fit the desired allocation.

There are also investors who aim to imitate the top-down approach adopted with public equity, whereby past performance of the assets in terms of risk, return, and correlation measures is used. These investors ultimately hope to exploit perceived low levels of correlation in different PE markets. While such an approach certainly has its merits, there are practical problems associated with it, mainly in terms of the limited availability and quality of financial data for private equity. As data become reliable only for fully liquidated funds, any quantitative approach based on historical data risks reflecting an outdated economic environment and will likely be even less indicative of the future. Even without being overly rigid or quantitative, a top-down approach can be used to generate alpha (by trying to spot the next big thing) or provide a sanity check that helps avoid fads. Relying on prescreening criteria derived from this high-level allocation, investors can limit the time they spend on analyzing individual investment proposals.

Apart from the questions associated with determining the weight of subportfolios, the major shortcoming of a top-down approach is that, in reality, strict allocations are not possible. In practice, it may be difficult to find and have access to a sufficient number of superior managers to fill each predetermined subclass allocation. Indeed, often only one or two superior managers operate in a particular sector, and each raises capital only every three to four years. On the other side, some superior managers may operate in a market segment without any allocation, thereby missing some potential high returns. Therefore, in order to adopt a top-down approach, investors cannot just wait for investment opportunities to arise but must proactively search for them so as to have a sufficiently large number of funds available to fill their desired allocations.

11.3.4 Mixed Approach

As both pure[5] bottom-up and pure top-down approaches are not problem-free, most investors follow a combined, or mixed, approach (see Exhibit 11.6).[6] Even a strong believer in the top-down approach would rarely invest in funds that are known to be of inferior quality just to fulfill a target allocation. Likewise, it may not be prudent for fund pickers to commit all of their money to a single sector based solely on the opportunity to invest in outstanding teams. Investors are conscious of the importance of diversification, but instead of diversifying on the basis of the correlations among the different asset classes, they define their target allocations on the basis of the investment strategies of the funds in which they invest.

Shearburn and Griffiths (2002) describe a modus operandi that can be considered representative of that followed by many PE funds of funds. They invest exclusively in established PE markets, such as the United States, the United Kingdom, and selected European economies; target the premier league of PE funds in Europe and the United States; and diversify by stage, focus, geography, and vintage year. The goal is to create a portfolio that is diversified according to specified investment strategies

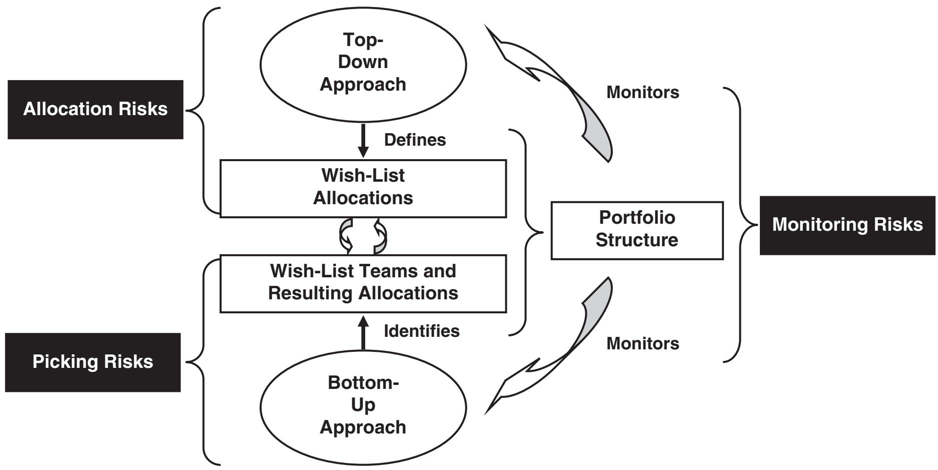


EXHIBIT 11.6 Mixed Approach

or dimensions. In this portfolio, all the strategies or dimensions have equal weight. More specifically, these authors describe an approach that "consists in creating a portfolio of unique PE strategies that are diversified from one another over multiple dimensions. These dimensions might include industry focus, investment size, geographic focus, and PE sub-asset class (such as leveraged buyouts, venture capital, growth capital and distressed investments)." They then assemble a portfolio of superior managers capable of generating extraordinary returns, with each manager's strategy being as distinct (i.e., uncorrelated) as possible from all the other strategies in the portfolio. Investors using this approach weight each strategy equally in the portfolio to minimize the concentration of funds. This means, for example, investing the same amount in a London-based large buyout PE manager as in a Silicon Valley–based early-stage venture capitalist. This manager-driven, equally weighted approach to portfolio construction also drives the weighting of PE subsectors, despite situations in which an investor may have a strong point of view.

Finally, there is an evolution in the adoption of the different approaches. For the early stages of an investment program, one of the main objectives is to quickly put capital to work with the best available funds in order to minimize performance drag resulting from idle liquidity. Young programs often cannot count on an established relationship, so the available universe of investment opportunities is likely restricted to a limited number of accessible wish-list funds. This makes a top-down approach difficult to implement, as allocation targets would make little sense. Once a sizable portfolio has been built, however, a top-down approach becomes a more appropriate means of identifying concentration (e.g., in sectors or in teams).

11.4 RISK-RETURN MANAGEMENT

The goal of portfolio design is to combine assets that behave in fundamentally different fashions in order to achieve a target risk and return profile. Because many of the traditional analytical approaches are not fully relevant to private equity, additional techniques are required. For example, by dividing the portfolio into two or more sub-portfolios, the management of the risk-return relationship can be improved. Similarly, understanding the benefits and limitations of diversification for portfolios invested in PE funds can lead to improved risk-return management.

11.4.1 Core-Satellite Approach

Broadly speaking, the **core-satellite approach** seeks to merge passive investing with active management in an attempt to outperform a benchmark. Under this approach, portfolios are constructed to minimize costs, tax liability, and/or volatility while providing an opportunity to outperform the benchmark. Typically, the core of the portfolio consists of passive and often increasingly low-cost investments that track the overall performance of an asset class. Additional positions, known as satellites, are added to the portfolio in the form of actively managed and higher-cost investments. The core is related to an investor's strategic asset allocation, while the satellite component is an effort to add alpha through tactical asset allocation.

More specifically, in the context of PE, the core-satellite approach can be explained as follows:

- The **core portfolio** typically aims to exploit established relationships, with institutional-quality fund managers raising funds that are expected to generate a predictable base return. If there is no change in the current environment, such mainstream funds are perceived to be the safe bet. A solid core of high-quality relationships allows LPs to stay in the game long enough to seize the golden opportunities, but exclusive reliance on the core would leave them susceptible to a long-term decline.
- The **satellite portfolio** can be interpreted as a bet on radical changes, which aims to explore new relationships (or opportunities). Typically, allocations to funds in the satellite portfolio are comparatively small and will, for the most part, have only a limited impact on the portfolio; once established, however, they can offer a more predictable base return and, over time, form part of the core portfolio. These funds should be seen as real options, having no value if not exercised (i.e., investing larger amounts in successor funds) from time to time.

This means that the portfolio is structured in various sub-portfolios, which can then be designed using one of the construction techniques (bottom-up, top-down, or mixed), and the portfolios can be constructed as layered pyramids (see Exhibit 11.7). For example, a well-diversified core, or bottom layer, may provide downside protection for the portfolio (risk aversion), while a less diversified satellite, or top layer, may look to generate upside gains (risk seeking).[7] This approach aims to increase risk control, reduce costs, and add value.

This may be an effective strategy for institutions that want to diversify their portfolios without giving up the potential for higher returns generated by selected active management strategies. Another advantage is the flexibility to customize a portfolio to meet specific investment objectives and preferences. The core-satellite approach also provides the framework for targeting and controlling those areas in which investors believe they are better able to control risks or are simply willing to take more risks. What constitutes core versus satellite depends on the investor's

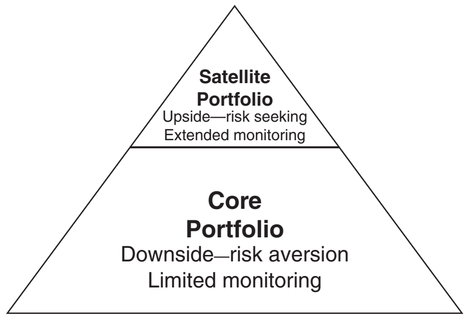


EXHIBIT 11.7 Core-Satellite Pyramid

focus and expertise. Some see venture capital as satellite, while others view a core portfolio as being balanced between buyout and VC funds. Another benefit to this approach is that it facilitates concentrating more effort on the satellite portfolio, which is expected to generate excess performance by incurring additional risk, and less time on the more straightforward, benchmarked lower-risk core portfolio.

To determine the balance between core and satellite (exploration) portfolios, several factors need to be taken into consideration:

- *The time horizon for the PE fund's investment program.* The more long-term oriented, the higher the value of the option and therefore the higher the degree of exploration that should be undertaken (i.e., bigger satellite).
- *The resources available.* With a larger reserve buffer, a higher degree of exploration is possible. The initial stages of a PE fund's investment program are primarily dominated by a core-based approach to build up slack resources that enable future exploration.
- *The changes (expected or unexpected) of the PE market environment.* The more innovatively disruptive a market environment, the more one needs to spread out one's options, whereas in a stable environment, exploration can be reduced to a minimum.

Layered portfolios as suggested by behavioral finance are sensible if not perfect ways, according to a Markowitz worldview, to allocate portfolio assets in private equity. The core-satellite approach provides a framework for targeting and controlling those areas where investors believe they are able to better control risks or are willing to take on more risk. This may be an effective strategy for institutions that want to diversify their portfolios in lower-cost investments without giving up the potential for outperformance offered by selected active management strategies.

11.4.2 Diversification

Diversification is another area in which LPs attempt to manage the risk-return relationship in PE funds. Diversification should be increased as long as the marginal benefit of adding a new asset to a portfolio (i.e., the marginal contribution of the asset toward portfolio risk-return optimization) exceeds the marginal contributions of competing asset classes (e.g., the source of funding) as well as the marginal cost (i.e., the transaction and administrative costs associated with making the allocation). Therefore, a key question related to the portfolio design is the optimal number of positions. For PE funds, it is difficult to determine the optimum diversification level. To resolve this, a series of additional questions should be addressed:

- What is the LP investor's ability to identify and, even more important, access top general partner (GP) management teams?
- What trade-off is the investor searching for between risk taking and profit seeking? In other words, what is the investor's risk appetite?
- Does the investor have objectives unrelated to the merits of the specific investment being evaluated?

Empirical evidence shows that the distribution of PE fund returns is quite dissimilar to the normal distribution.[8] Compared to the normal distribution, actual PE return distributions tend to exhibit higher probability of small losses and higher probability of very large positive outcomes. This means that the distribution has large positive skewness (see Exhibit 11.8). This situation arises because many PE investment strategies involve relatively small investments, with only a few having the potential to generate spectacular returns. The exception to this is the leveraged buyout (LBO) strategy, which can involve some relatively large investments, and although a large portion of an LBO investment is financed with debt, the equity portion can be relatively large compared to VC investments.

Diversification has two notable impacts on PE portfolios: (1) It lowers risk as long as asset returns are not perfectly correlated, and (2) due to the law of large

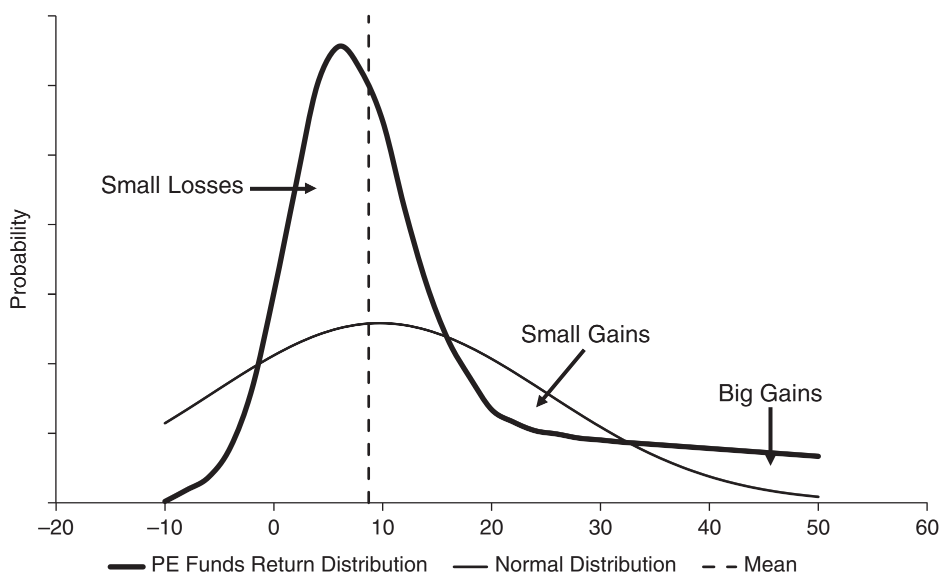


EXHIBIT 11.8 Private Equity Funds Return Distribution versus Normal Distribution

numbers, it brings the return distribution closer to the normal distribution. For PE funds, and especially VC, the distribution of returns shares characteristics with lotteries: A few extraordinary winners will compensate for many small losses. As in a lottery, in which buying all tickets guarantees picking the winner but also ensures that associated costs exceed the total gain, being invested in too many funds ensures that the few top performers likely do not adequately compensate for the many funds with mediocre or substandard returns. In addition, a significant amount of time and resources must be spent to identify skilled managers who would be willing to give access to the investor. Therefore, in VC, a common strategy is to make few bets under the assumption that selection skills (or luck) will deliver the winners without having to spend significant resources to manage and monitor a large portfolio that is bound to contain many losers.

There is no formulaic answer to the diversification question, but research suggests that for most assets, sufficient diversification is achieved with about 20 positions in whatever one is seeking to diversify.[9] Exhibit 11.9 illustrates the situation for U.S. VC fund portfolios, but the conclusions are similar for buyout funds and in Europe:

- About 80% of the standard deviation is diversified away with a portfolio of 20 to 30 funds.
- Skewness decreases more or less at the same rate as standard deviation.
- More than 80% of the kurtosis is diversified away with a portfolio of five funds.

The simulations performed in Exhibit 11.9 suggest that there is no advantage from a diversification perspective to having more than 20 to 30 funds in a portfolio. In fact, when investors are seeking diversification that will not eliminate the desired skewness and kurtosis, it does not make sense to have more than five funds in a portfolio.

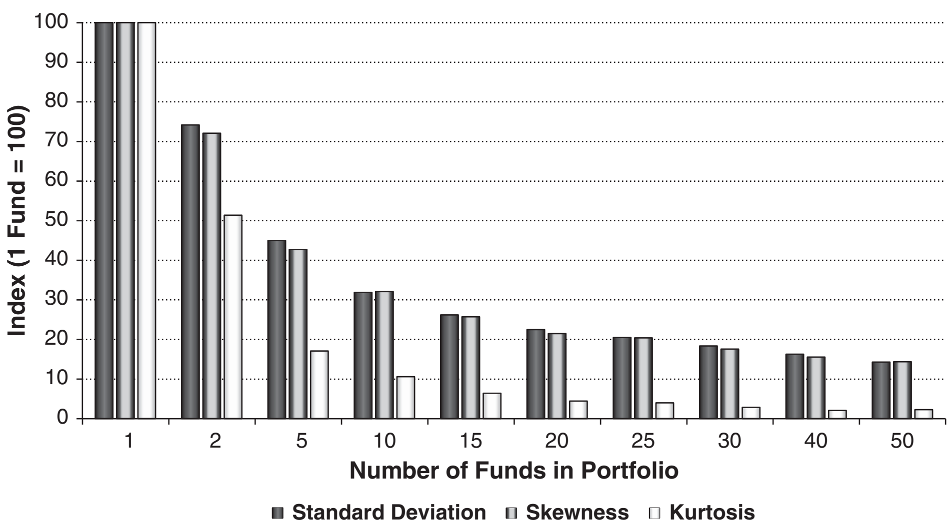


EXHIBIT 11.9 Standard Deviation, Skewness, and Kurtosis of Portfolio Returns with Increasing Diversification

Probably the strongest argument against a high level of diversification is the quick fading of expected fund quality. There are simply too few excellent fund management teams within a vintage-year peer group. Therefore, overdiversification not only leads to a reduction in expected positive skewness and kurtosis but also depresses the portfolio's expected return. The likely fading of quality and the expected loss of positive skewness and kurtosis are more of an issue for VC than for buyouts. For buyouts, the distribution function looks more symmetrical, and there are usually more high-quality teams. Finally, these notions hold only for the plain-vanilla limited partnership stakes in funds. More complex partnerships provide their investors with additional benefits, such as the option to coinvest and gain access to industry talents. In these cases, a higher level of diversification may be justified.

11.4.3 Naïve Diversification

Lhabitant and Learned (2002) argue that MPT is seldom applied to the full degree,[10] and that **naïve diversification** (also called 1/*N* heuristics) in practice usually produces reasonably diversified portfolios that are somewhat surprisingly close to the efficient frontier. In fact, naïve diversification is the optimal strategy when there is no information that allows differentiation among assets. However, naïve diversification approaches can be refined. Brands and Gallagher (2003) found that for private equity funds of funds, allocation approaches that ensure roughly equal allocations across investment styles perform better than portfolios constructed using randomly selected managers.

Albeit less rigorous than traditional portfolio models, naïve diversification is valuable to the prudent investor, as it can avoid extreme concentrations by ensuring an even distribution among the following dimensions:[11]

- Number of fund managers and stage focus
- Vintage years and calendar years
- Geography
- Industry sectors
- Investment styles and strategies employed
- Risk and fundamental metrics of portfolio companies

In general, a diversification strategy that doesn't take into account the specificities of this asset class can be quite inefficient. Overdiversification may lead to capping the upside. Furthermore, investing in many teams without managing the diversification of each risk dimension—such as industry sectors, geography, and, most important, the vintage-year spread—can seriously damage the portfolio. Investing in 20 funds will not lead to a meaningful diversification if all the funds are from the same vintage-year cohort or invested in companies with similar exposures (e.g., from the same industry).

There are also diseconomies of scale. The number of investments, rather than the invested amounts, sets the cost base (including legal expenses, due diligence, and monitoring efforts) of a portfolio of funds. It also becomes increasingly difficult to identify and gain access to suitable funds, as the number of quality opportunities is limited. Consequently, for PE fund portfolios, diversification may add costs in the management of risk versus return. On the flip side, having even smallish exposures to many GPs effectively serves as a call option for follow-on funds in cases where solid investment managers can be identified up front.

11.4.4 Cost-Averaging and Market-Timing Approaches

A proven method to achieve a consistent level of vintage-year diversification is to invest a fixed amount of money throughout all years and to steadily commit to the best funds around, regardless of whether the current environment is seen as good or bad. This disciplined approach helps minimize overexposure to vintage years with high valuations (or an unfavorable exit environment). It should be noted, however, that as size and quality of available opportunities are affected by the economic environment, it may not be feasible to allocate the same amount regardless of the economic environment, especially if the allocation is relatively large. Notwithstanding this potential problem, this systematic approach enables investors to build a diversified portfolio at average valuation. This **cost-averaging** approach sets an annual investment target for each PE fund type, or a more or less constant target throughout funding cycles, thus avoiding any inclination to try to time the cycles.

The **market-timing approach** involves varying investment levels across vintage years in an effort to invest more in years with better prospects and less in years with inferior prospects. When using this approach, it can be dangerous to assume that what has worked well in the past will continue to do so; this behavioral bias, referred to as **recency bias**, is the tendency to think that trends and patterns we observe in the recent past will continue in the future. Also, the PE market overreacts and experiences disruptions in trends. During market upswings, it is important to focus on quality and be more restrictive than the average investor is; during depressed market conditions, however, it could make sense to be more flexible, as very often the overall vintage-year quality can turn out to be attractive. It is important to stick to the budgeted allocation, as new opportunities to invest in VC or buyout funds tend to coincide with the rise and fall of returns of each strategy.[12] Successful investors often avoid the temptation to overweight commitments to the hot strategy of the moment; instead, they stick to a thoughtful and well-developed long-term capital allocation plan.

11.5 CONCLUSION

Private equity portfolios present investors with unique challenges. While there have been major advances in quantitative approaches to the construction of optimal portfolios for publicly traded securities, most of these developments have limited applicability for PE portfolios. To apply these techniques, investors typically assume that historical properties of various asset classes can be used to estimate those properties in future periods. For instance, correlations of various asset classes are estimated using historical returns, and these estimates are then used to determine the composition of optimal portfolios. The difference with PE funds is that their historical returns are of more limited use than other major asset classes in forecasting future outcomes. For example, the vintage year of a PE fund has been shown to be an important determinant of the fund's performance. This means that returns on previous vintage years cannot be used to estimate properties of future vintage years. In addition, the performance of PE funds is highly affected by the skills of the GP. Since the length of time between start and exit for each partnership is likely to be around six to eight years through its active life, there will be limited data on each manager.

All of these limitations mean that PE investors have to use rather simple and sometimes ad hoc methods to create their portfolios, which are likely to deviate significantly from optimal portfolios that can be constructed using the tools of modern portfolio theory. Still, some of the basic lessons of MPT can be employed. For example, we know that diversification benefits are maximized when asset returns are not highly correlated. In this context, investors should attempt to combine PE funds lacking common attributes associated with driving ultimate investment performance. That is, the funds should ideally differ in terms of industry, style/strategy, stage, leverage, vintage year, and location.

NOTES

1. We acknowledge the coauthorship of Thomas Meyer and Pierre-Yves Mathonet.
2. For a discussion of risk versus uncertainty, see Guerron-Quintana (2012).
3. The classification of the various private equity market segments should be driven by their different risk profiles. The classical ones are identified by their stage focus (venture capital vs. buyout) and geography (United States, European Union, and Asia), but more refined classifications are also used, including the less developed (but often important for portfolio design) emerging niche markets.
4. See Mathonet and Meyer (2007, chap. 4), for a detailed description and analysis of a real-world example.
5. Also, see Jaeggi (2005).
6. For more details, see Mathonet and Meyer (2007).
7. See Statman (2002): "The desire to avoid poverty gives way to the desire for riches. Some investors fill the uppermost layers with the few stocks of an undiversified portfolio like private individuals buy lottery tickets. Neither lottery buying nor undiversified portfolios are consistent with mean-variance portfolio theory but both are consistent with behavioral portfolio theory."
8. We are more likely to encounter symmetrical return distributions where there is equilibrium between supply and demand and many market participants push prices in one or the other direction. Therefore, it is not surprising that for PE return distributions are asymmetric.
9. See Flag Venture Management (2001), Weidig and Mathonet (2004), and Meyer and Mathonet (2005).
10. See Lhabitant and Learned (2002): "Very few investors effectively take correlations (that is, the non-linearity of risk) into account when making complex portfolio decisions. Rather, they prefer to allocate assets using simpler rules, such as dividing allocations evenly among the assets available. Many respondents even admitted to having no asset-allocation strategy at all!"
11. Standard & Poor's conducts ratings for structured notes backed by a portfolio of PE funds. The basis of the rating is the analysis of a portfolio of PE funds' diversification level in a multidimensional fashion. The following dimensions signal how diversified a portfolio is: number of fund managers or general partners and number of funds or limited partnerships, vintage years and calendar years, type of PE funds or strategies, industry or sector, geography, and single investment exposure. In the opinion of Standard & Poor's, it's mainly portfolio diversification that provides downside protection to holders of the rated notes (Erturk, Cheung, and Fong, 2001).
12. On this subject, see also Kaplan and Strömberg (2008), who note an inverse correlation between the amount of funds raised in a given vintage year and subsequent returns.

REFERENCES

Bain & Company. 2015. *Global Private Equity Report 2015*. Boston: Bain & Company.

Brands, S., and D. Gallagher. 2003. "Portfolio Selection, Diversification and Funds-of-Funds." School of Banking and Finance, University of New South Wales, Sydney, Australia.

Cambridge Associates. 2014a. "Global ex U.S. Private Equity & Venture Capital Index and Selected Benchmark Statistics, December 31, 2014."

———. 2014b. "Global ex U.S. Private Equity Index and Venture Capital Selected Benchmark Statistics, December 31, 2014."

———. 2014c. "U.S. Venture Capital Index and Venture Capital Selected Benchmark Statistics, December 31, 2014."

Cornelius, P. 2011. *International Investments in Private Equity*. Burlington, MA: Academic Press.

Erturk, E., L. Cheung, and W. Fong. 2001. *Private Equity Fund-of-Funds: Overview and Rating Criteria*. New York: Standard & Poor's Research.

EVCA (European Private Equity and Venture Capital Association). 2014. "2013 Pan-European Private Equity Performance Benchmarks Study." EVCA, Brussels, June.

Flag Venture Management. 2001. *The Right Level of Diversification*. Stamford, CT: Flag Venture Management.

Fraser-Sampson, G. 2006. *Multi Asset Class Investment Strategy*. Hoboken, NJ: John Wiley & Sons.

Guerron-Quintana, P. 2012. "Risk and Uncertainty." *Business Review* Q1:10–18.

Harris, R. S., T. Jenkinson, and S. N. Kaplan. 2015, June 15. "How Do Private Equity Investments Perform Compared to Public Equity?" Forthcoming in *Journal of Investment Management*. papers.ssrn.com/sol3/papers.cfm?abstract_id=2597259.

Jaeggi, A.P. 2005. "Successful Risk Management for LP's Private Equity Portfolios". Presentation, Super Investor Conference, November 17.

Kaplan, S., and P. Strömberg. 2008. "Leveraged Buyouts and Private Equity." NBER Working Paper 14207, July.

King, D., and M. Young. 1994. "Why Diversification Doesn't Work." *Real Estate Review* 25 (2): 6–12.

Lhabitant, F., and M. Learned. 2002. "Hedge Fund Diversification: How Much Is Enough?" *Journal of Alternative Investments* 5 (3): 23–49.

Mathonet, P., and T. Meyer. 2007. *J-Curve Exposure: Managing a Portfolio of Venture Capital and Private Equity Funds*. Hoboken, NJ: John Wiley & Sons.

Meyer, T., and P. Mathonet. 2005. *Beyond the J-Curve: Managing a Portfolio of Venture Capital and Private Equity Funds*. Hoboken, NJ: John Wiley & Sons.

Shearburn, J., and B. Griffiths. 2002. "Private Equity Building Blocks." *Pension Week* in association with Goldman Sachs, April.

Statman, M. 2002. "How Much Diversification Is Enough?" Leavey School of Business, Santa Clara University, September.

Swensen, D. 2000. *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*. New York: Free Press.

Weidig, T., and P. Mathonet. 2004. "The Risk Profiles of Private Equity." EVCA, Brussels, January.

CHAPTER 12

Measuring Private Equity Risk

Risk measurement for private equity is still in its infancy, and as Bongaerts and Charlier (2006) found, the academic literature on the intersection of private equity and risk management is very close to an empty set. At the time of this writing and despite recent regulatory initiatives, such as Basel III, Solvency II for the European insurance industry, and the Alternative Investment Fund Managers Directive (AIFMD), the situation has not significantly changed. Some early steps toward becoming more systematic about private equity risk measurement have been undertaken in recent years.[1] However, risks are typically still modeled in an overly simplistic and thus often misleading fashion, along the lines of risk measurement for listed assets, mimicking attempts to determine the standard deviation for periodic returns of stock price changes.

12.1 FOUR SIGNIFICANT RISKS OF PRIVATE EQUITY

Private equity has specific characteristics that are not shared by other asset classes. As a result, the standard risk management techniques that are used for other investments will not be effective. Investors should be aware of risks that play a far more important role in the performance of private equity investments. This chapter discusses the most significant risks in private equity, and examines risk management tools that are likely to be most effective in managing these risks. The most significant risks in this asset class are market risk, liquidity risk, commitment or funding risk, and capital or realization risk (EVCA 2013).

1. Market risk: **Market risk** is economic uncertainty with regard to the estimation and establishment of a price for an illiquid asset. As an illiquid asset class, the treatment of market risk in private equity poses conceptual challenges. There are two principal methods for valuing an asset. The first is its current market valuation (i.e., price), or an estimate of what that might be. The second is the present value of the estimated future cash flows from that asset. For publicly traded assets, in contrast to private equity, where there is no market price, market risk's impact on price and value is the same.
2. Liquidity risk: Limited partners can sell their stakes in private equity partnerships to fund their outstanding commitments. However, the secondary market for private equity investments is relatively small and highly inefficient. The characteristics of the secondary market expose investors to asset liquidity risk.

Moreover, secondary market prices are often influenced by factors beyond the fair value of the partnership, which often means prices are discounted. Liquidity risk plays a far less important role for public equities and only when emerging markets and small-capitalization stocks of developed markets are involved.

3. Commitment or funding risk: The unpredictable timing of cash flows over the life of a fund poses funding risk for the limited partner. Fund managers call most or all of the committed capital over the investment period of the fund. Limited partners then have to meet their commitments within a fixed short-notice period. Because commitments are contractually binding, a limited partner who cannot meet his obligations is forced to default on payments and lose a substantial portion of his share in the partnership. Funding risk is essentially absent in public equities, as current investments do not create an obligation to make future allocation to equities.
4. Capital or realization risk: In addition to the risk of losing invested capital due to liquidity constraints, private equity investors face the long-term risk of not recovering the value of their invested capital at realization. This long-term capital risk can be affected by a number of factors, namely (1) the ability of managers to create value and extract cash from investee companies, and (2) the level of equity markets at the time of exit. However, managers have full discretion as to the timing of divestments from investee companies within the lifetime of the fund and can wait for better or acceptable market conditions to exit their investments. Public equities have exposure to equity markets, but timing of exit does not play a significant role in their risk profiles.

Underestimating risk exposure of the economic capital gives the appearance that an investment in private equity is attractive and can result in overly risky commitments that could destroy capital, whereas overestimating risk exposure might lead to disregarding investment opportunities that appear unexciting but, in fact, could increase economic value.

Regulation broadly differentiates between operational risk, which covers legal and reputational as well as environmental, social, and governance (ESG)–related issues, and financial risk.[2] Private equity risk tends to be looked at mainly through the prism of operational risk, but questions surrounding operational risk are not specific to private equity. Regarding measurability, no conceptual problems need to be overcome: techniques exist and are fully accepted by investors and regulators alike. For the purpose of this discussion, **private equity financial risk** is considered to comprise mainly market risk, the risk of losses caused by adverse movements in market prices, and liquidity risk. Liquidity (that is, the ease and speed with which an asset can be sold at a fair price) is an important factor in an allocation decision and, as a consequence, constrains the allocation to illiquid assets.

This chapter is not about risk avoidance or operational risk. There is, of course, interdependence among all these risks, but financial risk exists even if everything is in hand operationally. Moreover, operational risk to a large degree is independent between different private equity firms and can therefore be efficiently diversified away by investing in portfolios of funds. It is, rather, financial risk that drives the extreme outcomes (particularly the upside) of private equity investments.

12.2 MODELING PRIVATE EQUITY

A crucial factor for a satisfactory modeling of private equity is that these investments are held only temporarily, with a preplanned or contractually fixed maximum term, which could be referred to as a value creation plan, and with the intention of profitable divestment.

12.2.1 Buy-to-Sell versus Buy-to-Keep

The key to private equity investing is its buy-to-sell philosophy, with a strong pressure to perform and reach the milestones targeted under a set value creation plan. Some institutional investors, such as sovereign wealth funds, pension funds, and corporations, follow the private equity modus operandi and implement a value creation plan, but then, after the restructuring, an indefinite holding phase takes place rather than a sell-off. Under this buy-to-keep assumption, private equity can be embedded in a portfolio structure together with publicly quoted equity. Unlisted companies may need to be given more time to be acquired or sold off but should otherwise follow the same value development and be subject to the same market influences as listed companies.

This is not the case during the implementation of a value creation plan under a buy-to-sell modus operandi, in which companies are restructured within a framework that is not subject to the same set of reporting requirements and regulations they were subject to before. Under this assumption, it is questionable whether a market value exists for a private-equity-backed company in isolation while the value creation plan is being implemented. Here, a risk model needs to take into consideration investing by the private equity firm, which (at least during the early stage of the intervention process) is inseparable from the portfolio company. Only after the restructuring is completed can the portfolio company be viewed as an independent entity that is exposed to market risk again.

12.2.2 Private Equity as Arbitrage

Private equity investing may be viewed as attempting to benefit in three ways. First, the deployment of private equity capital is a search for opportunities in uncertain, underresearched, or overlooked niches, where information is proprietary and where there is little or no competition. A second potential arbitrage opportunity is based on benefits from restructurings wherein portfolio companies' structures and corporate governance are moved toward the generally accepted shareholder model.

Third, and in line with this arbitrage model, some, such as those at the Yale endowment, argue that investors routinely overpay for liquidity and that inefficiencies in illiquid markets offer rich pickings for astute investors.[3] As Cornelius et al. (2013) explain, it is precisely this illiquidity risk, and more specifically the associated risk premium, that attracts some investors. Also, for this reason, the organized private equity market is dominated by funds that are structured as limited partnerships as the principal financial intermediary. The very features of this structure are essential for reaping the illiquidity risk premium that private equity offers to investors. In fact, for this reason, funds are structurally illiquid: in contrast to asset classes that may become illiquid thanks to financial turmoil and heightened risk aversion, limited partners are aware ex ante of the risk they take.

Not all investors are able to harvest this risk premium, however. As a matter of principle, only long-term investors can, since their liability profile allows them to lock in capital for a prolonged period of time, usually 10 years or even more.

12.2.3 How Risky Is Private Equity?

Markets are viewed as crucial for risk management, as they, in theory, provide liquidity to investors who can correct an investment error by selling the asset back to the market. Arguably, for marketable assets, market risk is the dominant risk, as liquidity is not an issue except during periods of extreme stress. Private equity funds, particularly, are illiquid, at least for risk management purposes. The secondary market for such assets is too opaque and not deep enough, transactions are too infrequent, and often sellers are acting under compulsion. As a route to liquidity, secondary markets tend to dry up precisely when liquidity is needed most. Therefore, funds need to be mainly understood as assets that are characterized by the statistical properties of their cash flows. After having committed to a fund, the LP has no control over timing and amount of cash flows.

For traded assets, market risk captures both dimensions, whereas for private equity risk, one needs to differentiate between capital risk (i.e., when invested capital plus a premium are not returned) and liquidity risk. From an LP's perspective, the main risks relate to the uncertainty regarding timing and amount of these cash flows (liquidity risk) and to what degree a fund in its setup and focus is in line with standard practices (related to capital risk). Occasionally, investors fail to properly model and manage their liquidity and, due to premature forced sell-offs on the secondary market, suffer damage that goes far beyond mere negative investment returns, giving an overall impression that private equity is a very risky asset class.

However, under the assumption that an investor remains committed over the fund's full lifetime, the capital risk of a diversified portfolio of funds is much lower than widely perceived.[4] Clearly, diversification is important. Research conducted by Weidig and Mathonet (2004) shows that for portfolios of 15 to 20 funds, there were surprisingly strong positive diversification effects.

In what way are companies that are bought-to-sell actually exposed to financial markets? In fact, it could be argued that private equity backers are shielding fledgling portfolio companies in their early stages (and those that are being restructured in turnaround situations) from disruptive market influences. Investors can provide portfolio companies with a lifeline despite adverse developments in the capital markets.

Finally, arbitrage is a low-risk activity, provided, of course, that there are suitable opportunities. Note that here, incomplete and poor-quality data are consistent with increased potential arbitrage opportunities, which is actually part of private equity's value proposition and puts the much-lamented shortcomings of industry statistics into perspective.

Diversification of portfolios of funds, the protection that funds provide to their portfolio companies, and private equity as arbitrage thriving on proprietary data are factors that change our perception regarding this asset class's financial risk, provided that liquidity risk is properly controlled.

12.3 WHAT IS THE VALUE OF A PRIVATE EQUITY ASSET?

How can we define the value of PE investments and determine how market fluctuations impact that value? There are two principal methods for valuing an asset. The first is its current market valuation as evidenced by prices observed in recent transactions, or an estimate of what that price might be. The second is the present value (PV) of the estimated future cash flows from that asset. Normally, liquidity and arbitrage in the market force these alternative methods of valuation into close alignment, but lack of liquidity and other market dysfunctionalities can cause the two approaches to diverge, occasionally sharply. Uniqueness of PE investment and even choice of portfolio companies are primary reasons why market forces may not bring prices close to the intrinsic value of the assets (PV of cash flows).

Much confusion regarding valuation in private equity comes from the way that value and price are habitually mixed up in the discussion. Valuations are subjective by definition, whereas prices are objective in the sense that they can be transacted upon. Despite the notions of fair value, asset prices for private equity either do not exist or are not observable.

12.3.1 Investments in Funds

In the context of private equity funds as the focus of institutional investing in this asset class, there are additional points that require reflection. According to EVCA Risk Measurement Guidelines, the relevant unit for risk assessment is the LP's share in the private equity fund, but this is not equivalent to the LP's share in the fund's portfolio companies reflected in the fund's net asset value (NAV).[5]

An LP has no option to initiate the sell-off of a portfolio company; this decision is entirely in the hands of the fund managers. Notably, there is no link between valuation changes in the fund's portfolio and the cash flows from and to its LPs; for instance, even if a portfolio company's value goes up significantly during a fund's investment period, it rarely results in an exit and thus a return of capital to the LPs. Valuation changes on the portfolio level and the resulting cash flows from and to the LPs are only indirectly related, with the probability of cash flows taking place dependent on the fund's stage in its life cycle.

From the risk manager's perspective, the fact that LPs are usually committing to blind pools causes another major headache. Initially, the fund holds no quantifiable assets, and the portfolio that will be reflected in the NAV builds up over time. However, LPs are immediately exposed to liquidity risk, which is close to its maximum on day 1, when the fund will call capital for management fees and also for the first investments.

12.3.2 Impact of Undrawn Commitments

When looking at limited partnership funds, the crucial question relates to undrawn commitments: Do they belong to the fund and thus become part of the allocation to private equity? While they are often viewed as irrelevant, undrawn commitments cause serious problems for LPs during financial crises.[6] This experience during the most recent crisis resulted in increased interest by practitioners and academics alike in what has been coined commitment risk or funding risk.

One aspect of this type of risk is overcommitments, in which investors commit more to funds than they have available as resources. In the case of overcommitments, future distributions from one fund are needed to honor the capital calls for another fund. The valuation of the fund's underlying assets is reconcilable with a present-value perspective for the fund only if the LP is not following an overcommitment strategy. In cases in which distributions are insufficient, the LP is taking on commitment risk. **Commitment risk** describes the situation in which an LP may become a defaulting investor if the proceeds of exiting funds are not sufficient to pay the capital calls of newly committed funds. This can also be unintentional; for instance, exchange rate changes may have an adverse impact on the resources available to respond to capital calls.

Otherwise, shortfalls stemming from funds that are not repaying capital in time will have an adverse impact on valuations of other funds for which capital calls cannot be responded to by the LPs. Deep-pocketed LPs usually view their undrawn commitments as irrelevant from a liquidity standpoint, given their relatively small allocations to illiquid partnerships compared with their holdings of public equity and bonds. Investors also believe that they can avoid commitment risk simply by not overcommitting and by holding liquidity in risk-free and extremely liquid instruments, such as Treasury bills. Holding undrawn assets in public equity may reduce the ability of LPs to fund capital calls during a market downturn.

The question is whether there are opportunity costs associated with the undrawn commitments locked up in liquid assets. According to finance's standard model, assets are valued by breaking them into various components.[7] From this perspective, it is often argued that Treasury bills are zero net present value (NPV) investments, with low returns reflecting the absence of risk; thus, there should be no opportunity costs. However, the value of such a fund before its first investment would be zero, as the fund does not yet hold any asset, or even negative, as the valuation needs to factor in the future management fees. One major problem obviously lies in the conventions of how funds are modeled.[8]

Financial theory is not questioned here, but the flaw appears to lie in the method of breaking down a fund into its components. Investments in funds are occasionally modeled as loans, with the LP acting as a lender. Cornelius et al. (2013) suggest the opposite view: economically, the situation is as if the GP called the entire amount of commitments on the first day of the fund's lifetime and lent the undrawn capital back to the LP. While this may seem counterintuitive, this approach is motivated by the observation that it is actually the GP who holds the power over the use of the undrawn commitments. As for an ordinary loan, the GP can restructure the relationship with the LP if the latter defaults. In this situation, the LP may lose the paid-in capital or suffer other penalties, with the GP holding the paid-in capital as collateral. There is also a clearly specified period of time during which the capital needs to be paid by the LP (i.e., the investment period).

Note that from this new perspective, the value of a fund at the time of commitment would be the commitment size, and it would be funded by debt (i.e., the undrawn commitments lent to the LP). Here, the GP is the owner of the undrawn capital, and it should not matter to whom the GP lends it. While the LPs happily, but probably mistakenly, accept a Treasury return for the undrawn capital they hold themselves, they would certainly not be willing to do so if the GP were to lend the undrawn capital to a third party. In this explanation, the pieces of the puzzle fall into place: funds of funds that, as LPs overcommit, aim to compensate for these opportunity costs essentially take on leverage. However, academic research into commitment risk is still in its infancy, so this view has not yet found general acceptance.[9]

12.4 APPLYING THE VAR CONCEPT TO PRIVATE EQUITY

The value at risk (VaR) concept has become part of the prudential apparatus of financial institutions and is based on simplistic forms of modeling and calculation. VaR was first allowed in regulatory capital modeling in 1996. For a long time, VaR has been the preferred measure of market risk for regulators, and concepts similar to VaR are used in various parts of financial regulation. VaR aims to estimate likely portfolio losses in extreme situations based on statistical analysis of historical price trends and movements.

12.4.1 Problems and Limitations

In finance, volatility is generally seen as the main measure for risk, underlying the VaR concept. In private equity, a widely accepted, albeit problematic, approach to calculating the risk of fund investments is to look at the NAV's volatility. In principle, this is done by looking at value changes in a portfolio's NAV time series, typically quarterly.[10] Volatility is estimated by measuring the returns reported at the different time intervals based on quoted indices or other benchmarks available in the market.

Arguably, volatility calculated in this way is mainly relevant for the buy-to-keep modus operandi of private equity investing. It is, however, not a meaningful measure of risk for buy-to-sell assets that follow a life cycle, such as the J-curve, and funds with distinct investment and disinvestment periods. The NAV time-series approach tries to project future development based on a relatively short history, which often does not allow for a determination of the trajectory of the funds' value development.

12.4.2 A VaR for Portfolios of Private Equity Funds?

As VaR was originally designed by investment banks heavily involved in trading in publicly quoted assets, one may ask whether this is the right approach for illiquid asset classes. However, extensions like credit VaR demonstrate that related risk measures can be derived in a similar fashion as those for market risk. How can a VaR approach be applied to the case of portfolios of private equity funds while reflecting their long-term characteristics?

For a given portfolio of funds, probability, and time horizon, the VaR at time t_0 is defined as a loss on the portfolio over a given time horizon $\left[t_0, t_1\right]$ and for a given confidence level α, $0 \le \alpha \le 1$. For calculating VaR, it is generally assumed that there are normal economic conditions and no changes in the composition of the portfolio of funds.[11]

12.5 CALCULATING VAR BASED ON CASH FLOW AT RISK

Financial institutions use VaR as a basis for capital adequacy and measuring traded risk, but nonfinancial firms have found this concept difficult to apply in their risk management because here value mainly takes the form of real investments in fixed

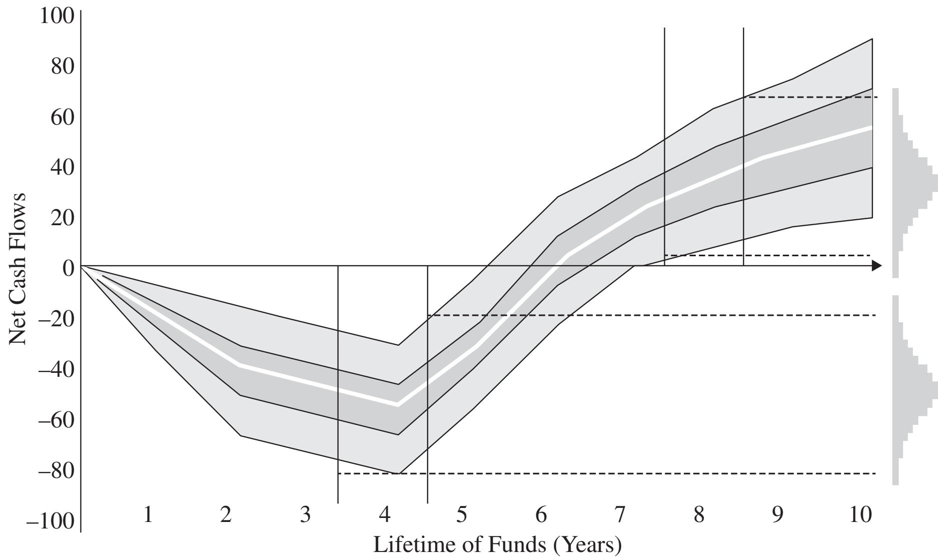


EXHIBIT 12.1 Cash Flow Distributions at Two Different Time Intervals to Determine CFaR

assets that cannot be monetized easily. Industrial companies tend to look at the cash flow at risk (CFaR) as a more relevant measure for their risk exposures. The **cash flow at risk** is the maximum deviation between actual cash flows and a set level (e.g., a budget figure) due to changes in the underlying risk factors within a given time period for a given confidence level.

While in the case of tradable assets VaR is usually computed for very short periods of time (days or weeks), CFaR relates to longer periods, typically quarters or even years (see Damodaran n.d.). In the case of financial firms, it is argued that marked-to-market portfolios are convertible into cash at short notice, and, therefore, their VaR is also their CFaR (see Damodaran n.d.; Yan, Hall, and Turner 2011). However, as discussed before, this argument is clearly not true for illiquid assets, such as private equity.

When looking at the CFaR for a portfolio of funds, we focus on variations in cash flow within a given time interval. For an LP, both directions of cash flows are relevant: positive cash flows, as they will be needed by the LP for new investments or to honor future capital calls; and negative cash flows, as they expose the LP to liquidity risks because available cash may not be sufficient to meet the LP's financial obligations (see Exhibit 12.1).

12.5.1 Value at Risk Based on Cash Flow Volatility

Instead of looking at a time series of returns, the idea is to take the so-called terminal wealth dispersion as a reference point for measuring risk, which relates directly to expected return and volatility levels. As market valuations are not regularly available for funds, the time period relevant for risk assessment is the entire lifetime of the portfolio of funds. Due to the short history of available statistics for such funds and their long-term nature, incorporating data for only fully liquidated funds would leave us with too few data points. Instead, practitioners also include mature funds that are still active but exceed a set threshold for the minimum age in the underlying data sample.[12]

Calculating their cash inflows and outflows, while reflecting the last reported NAV, results in the calculation of the total value to paid-in (TVPI) ratio. Based on the outcomes for the TVPI for each fund, a probability density function for the terminal wealth dispersion can be determined. For mature funds, the NAV has a lower weighting, as the investment period is already over and the first distributions and exits have already occurred. Therefore, this approach not only takes into account the changes of the NAV but also reflects the cash flow behavior.

The risk profile for a portfolio of funds is derived from the returns of comparable mature funds.[13] This avoids to a large degree issues related to too few data points, such as autocorrelation and desmoothing, but is based on the quite heroic but typical assumption that funds in the currently held portfolio will perform like funds in the past.

How is this VaR calculated? An investor at time t_0 wants to determine the VaR of the portfolio of funds held for some time t_1 in the future, typically the end of the year. To do this, the investor needs to determine the probability distribution for the portfolio of funds' terminal wealth dispersion at time $t_1 > t_0$. One way to approach this is by performing Monte Carlo simulations and randomly drawing out of a database returns for mature funds that reflect the characteristics of the portfolio to be modeled. There are eight steps:

1. The higher the number of runs for the Monte Carlo simulation, the more stable the results will be.
2. For every fund in the limited partner's portfolio, a specific cash flow scenario consistent with the fund's projected range for TVPIs and lifetimes is generated.[14]
3. To project cash flows in every simulation run, a new set of randomly chosen parameters is generated as inputs for the fund model.
4. Correlations are reflected by constraining the random draws to subsections of the database (e.g., specific vintage years or strategies).
5. All individual fund scenarios are aggregated, producing a scenario for the entire portfolio of funds' cash flow.
6. A suitable discount rate is applied to determine the PV for this portfolio scenario.
7. The PV results from each simulated portfolio scenario are compiled, resulting in the distribution function.
8. Based on this distribution function, the VaR for confidence level α is determined.

The yearly standard deviation of returns can be derived from the terminal wealth dispersion by essentially annualizing backwards (see Exhibit 12.2) to mitigate the misleading impact of interim valuations.

This analysis reflects the risks of an investor in a portfolio of funds who has enough liquidity to honor all capital calls and thus does not feel pressured to sell at unfavorable conditions during the portfolio's lifetime. This is a crucial assumption, as many large institutions find themselves in this situation, with only a small allocation to alternative assets and no liquidity issues, and reflects what they perceive to be the true risk of investment in private equity and real assets. In general, however, a funding test has to confirm that the LP is in such a position.[15]

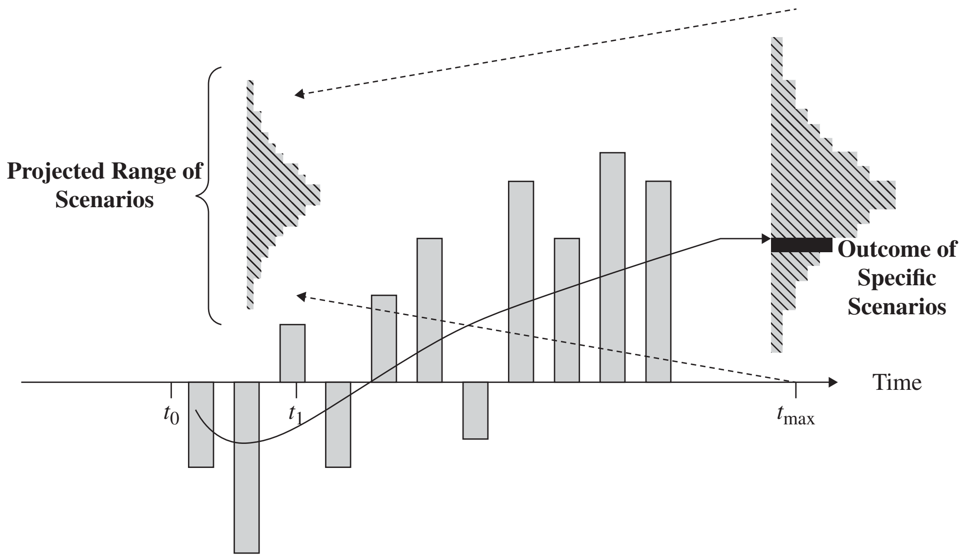


EXHIBIT 12.2 Annualizing Final Returns Backwards

12.5.2 Fund Growth Calculation

In a fund growth calculation, we begin by calculating the fair value of a fund at time $t = 0$ based on m simulations of cash flow series over the entire lifetime (n periods) of this fund. For each scenario, the straight-line growth over its full lifetime and the resulting gain or loss per time period are calculated. The VaR for a given time period is derived from projecting gains and losses under all scenarios starting from the fund's fair value at $t = 0$.

Here we look at not only one but a number of simulation paths for each fund over its remaining lifetime. In addition, the VaR is calculated based on the difference between the present value of each simulation run and its current value.

Through the additional scenarios, the higher uncertainty that is implicitly included in the approach should also have an influence on the result of the distribution of the density function, which is expected to be broader than in the time-series calculation described earlier. Moreover, this allows us to calculate, in theory, a VaR for a portfolio comprising one fund only. Here, calculating the VaR requires the following steps:

1. For each fund in the sample, generate cash flow scenarios over the fund's full lifetime.
2. For each scenario and a given discount rate, calculate PV_i.
3. Calculate the average present value of all scenarios of one fund in order to derive the fair value of the fund at time $t = 0$:

$$Avg(PV) = \frac{1}{m}\sum_{i=1}^{m} PV_i \tag{12.1}$$

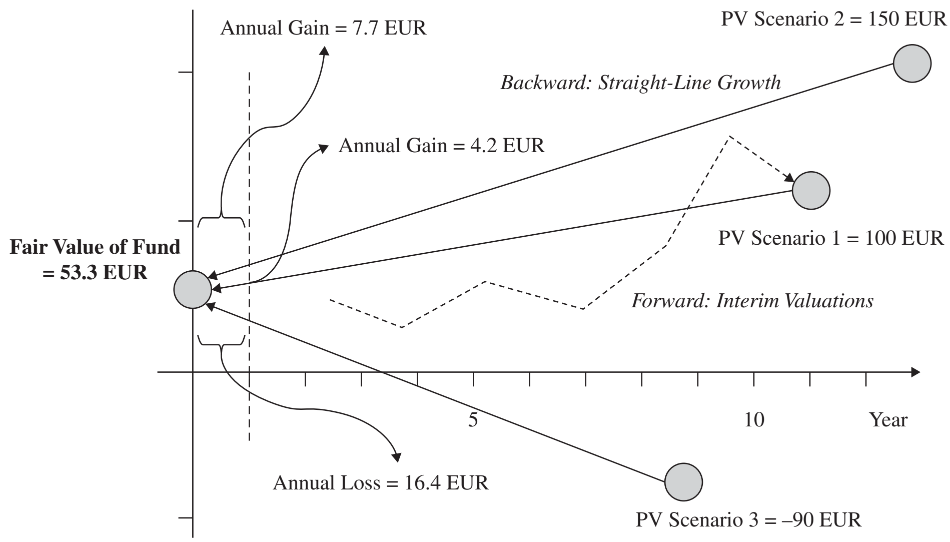


EXHIBIT 12.3 Fund Growth Calculation Approach

4. The period gain or loss for a given scenario is determined by relating the difference between the fund's fair value to the time period (depending on whether quarterly or annual VaR is being calculated). This linear approach eliminates the fund's J-curve or other distortions induced by the fund's life cycle:

$$GAIN/LOSS = \frac{Avg(PV) - PV_i}{n} \tag{12.2}$$

5. Based on these results, the density function of the quarterly or annual gains and losses can be computed. This allows us to calculate the portfolio's VaR over the required period for a set confidence level.

To understand this approach, consider the following simple example. Let us assume that we want to determine the VaR for year-end (i.e., after one year) and for just one fund.

For this one fund, we run three cash flow scenarios with a given discount rate, giving three PVs (see Exhibit 12.3):

SCENARIO 1: Fund's remaining expected lifetime is 11 years.

SCENARIO 2: Fund's remaining expected lifetime is 12.5 years.

SCENARIO 3: Fund's remaining expected lifetime is 8.75 years.

These three scenarios are based on different cash flow scenarios and, hence, reflect possible real outcomes of the fund. Assuming that they will materialize with equal probability, the fair value of the fund at time 0 by definition is the average of these scenarios:

$$\frac{150 + 100 - 90}{3} = 53.3\,\text{EUR} \tag{12.3}$$

We calculate the risk of this fund based on these scenarios and the gain/loss over a given time period. Under scenario 1, the fund would gain a value of 4.2 EUR per annum (p.a.) over its projected lifetime:

$$\frac{100 - 53.3}{11\ \text{years}} = 4.2\,\text{EUR p.a.} \tag{12.4}$$

Under scenario 2, the gain is 7.7 EUR p.a.; and under scenario 3, the fund would lose 16.4 EUR p.a. Based on these results, the density function for the fund's valuation after one year can be determined as follows:

1. Should scenario 1 materialize, after one year the fund would be worth 53.3 EUR + 4.2 EUR = 57.5 EUR.
2. Should scenario 2 materialize, after one year the fund would be worth 53.3 EUR + 7.7 EUR = 61.0 EUR.
3. Should scenario 3 materialize, after one year the fund would be worth 53.3 EUR − 16.4 EUR = 36.9 EUR.

With these valuations, the risk of losing any capital until the next year can be assessed. Based on these results, we can determine the histogram and density function for many scenarios of fund valuations at a future point in time. This forms the basis for determining the VaR over a required projection period (e.g., annually or quarterly).

12.6 CONCLUSION

The VaR concept has been controversial since it moved into the public eye during the mid-1990s. It has generally been criticized for a number of reasons, and academics see many conceptual problems with it. Much of the literature focuses on the VaR's failure to fully satisfy the axioms that academics have identified as necessary for a metric to be coherent.[16]

For most risk managers, on the other hand, the VaR's flaws are rarely relevant problems in practice, as managers should be less concerned about precision and complement the VaR by stress testing anyway. Despite various reservations and based on practical experience, it is believed that a VaR model is a useful tool as long as one is aware of its limitations. In any case, instead of relying on one magic formula, investors would be better advised to look at several risk indicators simultaneously, as no single metric will be able to give the full picture.

As previously argued, for investors in illiquid assets, CFaR is actually more relevant. When discussing VaR, the main point is to address regulatory requirements that ask to look at illiquid assets together with publicly quoted assets in a unified way. Whether it is meaningful to compare market prices and model-based valuations is another question, however. In the context of illiquid assets, the portfolio's CFaR should be reconcilable with its VaR, which is why Cornelius et al. (2013) advocate an approach based on cash flow volatility for calculating a VaR in such cases.

NOTES

1. See EVCA (2013) and Cornelius et al. (2013).
2. As far as Basel III is concerned, regulatory capital is calculated for three major components of risk that a bank faces: credit risk, operational risk, and market risk. Other risks are not considered to be fully quantifiable at this stage. Solvency II considers market, credit, liquidity, insurance, and operational risk, and the AIFMD views market, credit, and liquidity risk as financial risk.
3. See Plender (2014).
4. See Weidig and Mathonet (2004).
5. See EVCA (2013) and an in-depth discussion in Meyer and Mathonet (2005).
6. Cornelius et al. (2013) mention a number of examples, such as Harvard Management Corporation, which manages the endowment of Harvard University, and CalPERS, the largest U.S. pension fund.
7. From the LP's perspective, the stake in the fund cannot be broken up into individual portfolio companies as long as they are in the intervention phase.
8. Fraser-Sampson (2013) points out that these are "cash-flow type assets," and "while finance theory is happy to apply a different approach to the ones they like (bonds, that are measured using compound returns) they illogically deny it to the others, which include private equity, infrastructure and real estate as well as various energy type assets such as oil and gas royalties."
9. See Cornelius et al. (2013) and Meyer (2014).
10. See McCrystal and Chakravarty (2011).
11. While regulators require confidence levels of typically 99% to 99.5%, this is not truly meaningful for illiquid assets, with their limited number of data points.
12. For private equity, Diller and Herger (2008) define this as funds with an age of at least five years.
13. See Weidig and Mathonet (2004).
14. For an in-depth discussion of methodologies for determining a fund's TVPI and lifetime ranges, refer to Cornelius et al. (2013).
15. See EVCA (2013).
16. See Jorion (2001). A famous 1997 debate between Nassim Taleb and Philippe Jorion set out some of the major points of contention. See Jorion and Taleb (1997).

REFERENCES

Bongaerts, D., and E. Charlier. 2006. "Risk Management for LBOs in Buy-and-Hold Portfolios." August 27. Available at www.greta.it/credit/credit2006/poster/2_bongaersts_charlier.pdf.

Cornelius, P., C. Diller, D. Guennoc, and T. Meyer. 2013. *Mastering Illiquidity: Risk Management for Portfolios of Limited Partnership Funds.* Chichester, UK: John Wiley & Sons.

Damodaran, A. N.d. "Value at Risk (VaR)." http://people.stern.nyu.edu/adamodar/pdfiles/papers/VAR.pdf.

Diller, C., and I. Herger. 2008. "Private Equity—Will You Take the Risk?" Private Equity International, May, 106–9.

EVCA. 2013. *Risk Measurement Guidelines 2013*. European Private Equity and Venture Capital Association, January. www.evca.eu/uploadedfiles/EVCA_Risk_Measurement_Guidelines_January_2013.pdf.

Fraser-Sampson, G. 2013. *Intelligent Investing: A Guide to the Practical and Behavioural Aspects of Investment Strategy*. Basingstoke, UK: Palgrave Macmillan.

Jorion, P. 2001. *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill.

Jorion, P., and N. Taleb. 1997. "The Jorion–Taleb Debate." *Derivatives Strategy*, April.

McCrystal, A., and N. Chakravarty. 2011. "Solvency II: Private Equity, the LPX50 and Risk Calibration." Pantheon Ventures Paper, November.

Meyer, T. 2014. *Private Equity Unchained: Strategy Insights for the Institutional Investor*. Basingstoke, UK: Palgrave Macmillan.

Meyer, T., and P.-Y. Mathonet. 2005. *Beyond the J Curve*. Chichester, UK: John Wiley & Sons.

Plender, J. 2014. "There Is a History Lesson to Be Learnt from Yale Endowment." *Financial Times*, September 28. www.ft.com/intl/cms/s/0/fc223c38-4406-11e4-baa7-00144feabdc0.html#axzz3uzwImEjX.

Weidig, T., and P.-Y. Mathonet. 2004. "The Risk Profile of Private Equity." January. Available at http://ssrn.com/abstract=495482.

Yan, M., M. J. B. Hall, and P. Turner. 2011. "Estimating Liquidity Risk Using the Exposure-Based Cash-Flow-at-Risk Approach: An Application to the UK Banking Sector." Working Paper 2011-06, Loughborough University, School of Business and Economics.

CHAPTER 13

The Management of Liquidity

Investments in private equity (PE) funds have proved to be risky for a number of reasons, but principally because of the long duration of the exposure and the lack of liquidity. This chapter is concerned with illiquidity risk associated with investments in private equity.[1] As discussed in previous chapters, private equity has rather unique features and is subject to a number of risks, with illiquidity being one of the important ones.[2] As noted, private equity is an equity investment in nonlisted companies. As the companies are not traded on a secondary market like the shares of publicly listed companies, there is no market price available on a regular basis. Only if the company is sold to another investor can true market values be observed, but this typically happens only after a number of years.

13.1 IDENTIFYING ILLIQUIDITY AND MANAGING CASH FLOWS

Managing the liquidity of a private equity fund investment program needs to take into account the existing interdependencies among the overall investment strategy, the management of the undrawn capital, and the available resources and aspects of timing. It is a difficult task to put money to work efficiently while maintaining a balance in the portfolio composition and the quality of the individual fund investments.

Investors do not pay in all of their capital on the first day; rather, the money is drawn by the fund over time. This represents a specific risk for investors—that is, funding risk. If an investor is not able to pay the capital call in accordance with the terms of the partnership agreement, the investor defaults on the payment. In such a case, the investor might lose the entire investment and all the capital already paid into the fund. Many fund managers have strict rules in their limited partnership agreements in the case of a defaulting investor. Typically, the investor will lose the entire investment; in some cases, the investor still holds the liabilities. This strict mechanism is important for fund managers, as they need to have the highest possible security to fund the investments they would like to acquire. In addition to the risk of not being able to fulfill their own undrawn commitment, each investor can be adversely impacted as a result of other investors defaulting. Hence, liquidity and funding risks arising through unfunded commitments are an important element and need to be reflected in sophisticated risk management systems.

13.1.1 Four Advantages of Modeling Private Equity Cash Flows

Modeling the cash flows of private equity investments is an important part of the liquidity management process and potentially allows one to do the following:

1. Improve investment returns for the undrawn capital.
2. Increase the profit generated by the private equity allocation through overcommitment.
3. Calculate an economic value when a discount rate is available.
4. Monitor the cash flows and risk-return profiles of a portfolio of private equity funds.

Achieving a high total return for the overall investment program is a complex task that requires not only technical skills, particularly in quantitative modeling and financial engineering, but also a high degree of judgment and management discipline. There is no quick fix for this, and only a disciplined approach can deliver small improvements that eventually add up to a significant impact. As a result, it is likely to take many years before an investment program is able to reach sustainable levels of high total return and a stable long-term allocation.

13.1.2 Defining Illiquid Assets

Different asset classes are subject to different degrees of liquidity (see Exhibit 13.1). The World Economic Forum (2011) discusses the case for long-term investments and, in that context, presents different asset classes according to their degree of liquidity and the relevant time horizon for their investors.

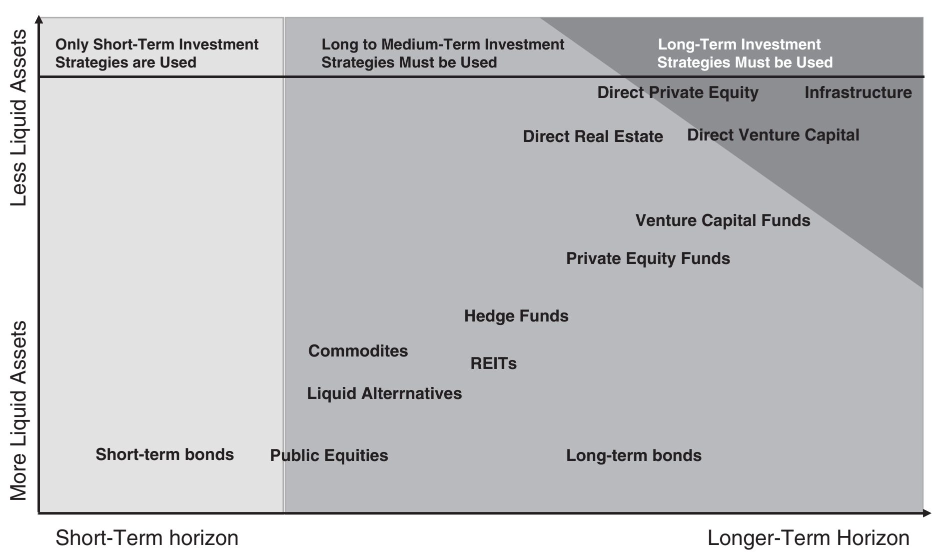


EXHIBIT 13.1 Asset-Class Liquidity versus Time Horizon

Private equity and real assets are structurally illiquid investments. **Structural illiquidity** is illiquidity inherent in the terms of an investment, such as a lockup provision. In the case of these investments, their low liquidity is due to the long ex ante lockup period of investors' capital. This sets structurally illiquid investments apart from other asset classes, whose degree of liquidity is generally high but ex post—that is, after the investment decision is made—may deteriorate sharply in periods of financial stress. Assets that are expected to be liquid but lose their liquidity during such periods are cyclically illiquid. Thus, **cyclical illiquidity** is illiquidity caused by market conditions such as financial stress. We have seen that during periods of financial stress, investments such as collateralized debt obligations, high-yield corporate bonds, and even money market funds may become illiquid. The primary reason for increased illiquidity of investments that are expected to be liquid is adverse selection, which arises when there is asymmetric information between sellers and buyers. When sellers have better information than buyers have, buyers can expect to see assets of below-average quality being offered for sale. Consequently, they would be reluctant to purchase those assets, leading to a lower level of liquidity in the market. Since the level of adverse selection is expected to be higher during periods of financial stress, liquidity tends to be lower during such periods.

13.1.3 Funding Risk as a Source of Illiquidity Risk

There are two sources of illiquidity risk for private equity investors: funding risk and exit risk.[3] **Funding risk**, also referred to as default risk within the private equity industry, is the risk that an investor will not be able to pay his capital commitments to a private equity fund in accordance with the terms of the obligation to do so. If this risk materializes, an investor can lose the full investment, including all paid-in capital, which is why it is of paramount importance for investors to manage their cash flows to meet their funding obligations effectively. The financial crisis in 2008 highlighted the importance of managing funding risk.

One possible reason for running into liquidity issues would be an investor following an overcommitment strategy. As private equity funds do not typically draw all of their committed capital, or as some companies have already been exited before all the commitment has to be paid in, the net liquidity need is typically smaller than the commitment size. Therefore, investors need to run an overcommitment strategy to avoid being permanently underexposed to their strategic allocation to the asset class. In case of market turmoil, however, an overcommitment strategy may result in severe issues for investors, as they may be required to pay out more money to meet commitment obligations than they had anticipated and/or than is available.

The second reason for running into liquidity issues is that investors may follow a self-funding strategy, whereby those investors who have been running a private equity portfolio for some years typically use the distributions of mature private equity funds to finance the capital calls of young funds. Depending on the maturity of the existing portfolio, a private equity program can be set up and managed in a way that the capital calls and distributions are in a steady state and can be matched accordingly. Therefore, no additional capital needs to be put into the private equity program, as it is self-financing in normal market conditions. Indeed, it is common for mature PE fund portfolios to face the problem of how to ensure that a desired level of exposure is maintained, as the level of distributions outstrips the level of new commitments being drawn, so the overall level of exposure to private equity declines. However, if a market distortion suddenly occurs and distributions are missing because exit activity on underlying companies dries up, investors may run into problems, as they would require additional capital from external sources to meet their commitments. Other external sources could be regular capital inflows from a main business (e.g., premiums collected by a pension fund); sale of liquid assets, such as investment-grade corporate bonds and listed equities; or sale of any other investment—potentially including their private equity funds—on the secondary market.

Theoretically, LPs have the option to abandon their investment simply by defaulting, in which case the exercise price of the option is the default penalty that should be weighted against the undrawn commitments that can be saved. It is argued that the threat of capital withdrawal is a useful contractual tool to reduce agency costs between investors and low-quality funds. However, the default involves a reputational penalty that investors suffer when they exercise their walk-away option. This penalty may be substantial, as a defaulting investor might not be allowed to invest in other PE funds.

Funding risk can be measured through a funding test or through cash flow models that take extreme cases into account. The funding test places the undrawn commitments in relation to the resources available for commitments. Alternatively, a cash flow model provides the investor with a simulation of the expected capital calls and distributions in the future. Investors can reduce the funding risk by assessing their future commitment plan with cash flow simulations and cautious planning. Investors who have limited external capital available or large allocations to illiquid assets should be more cautious on the overcommitment and/or self-funding strategy. However, when deciding on such a strategy, investors should be aware of possible extreme scenarios and how much cash would be necessary, as well as how this could be obtained from other sources. A strategic plan for these extreme cases as well as the portfolio construction plan is the key element.

13.1.4 Exit Risk as a Source of Illiquidity Risk

Exit risk is the risk that an investor is unable to redeem or liquidate his investment at the time of his choosing or would exit at depressed valuations. As we have already discussed, this source of illiquidity is a direct result of the structural liquidity of the underlying asset. Private equity fund structures are designed so that investors remain in the fund for the full term, without an opportunity to redeem their commitment. As a result of these structures, however, a secondary market for LP commitments has evolved. Consequently, liquidity risk may also be regarded as the risk that an investor wants to sell a private equity investment on the secondary market, but the market does not offer enough volume or efficiency for a fair trade. Moreover, secondary market prices are often significantly affected by factors unrelated to the fair value of the partnership, which result in prices being discounted. For instance, investors selling from a distressed position often have to accept discounts to reported net asset value (NAV). Liquidity risk in private equity is difficult to measure. While the secondary market can be very active in a normal market environment and during boom phases, this level of activity is far from what one would see in even the most illiquid of listed markets. Moreover, the secondary market was nearly shut down during the financial crisis in 2008–2009 with very low trading volumes. Therefore, the liquidity risk for investors in private equity seems to be high due to inefficient secondary markets. Liquidity risk in private equity is difficult to reduce, although it is simpler to handle for investors in an overall asset allocation model. If an investor is solely focused on private equity assets and needs to sell in difficult market times, the investor cannot circumvent the liquidity risk. However, if private equity is only a small part of a well-diversified asset allocation, as is the case for many institutional investors, many other assets are more liquid and can be traded.

As discussed, the secondary market may be exploited to provide liquidity. However, to some degree, the secondary market may be subject to what is known as the lemon problem, which is directly related to the adverse selection problem previously mentioned and exists because of asymmetric information between the buyer and the seller. In the secondary market, the buyer does not usually know beforehand the quality of the PE investment being offered. Therefore, because the buyer's best guess is that the private equity is of average quality, the buyer will be willing to pay the price of an average-quality PE fund. This means that the seller of a high-quality PE fund will be unable to get a high enough price to make the selling of the position worthwhile. As a result, the seller may be reluctant to sell the fund.

The same dynamic can be observed in all markets that are subject to asymmetric information. In asset markets, where there are doubts about asset quality, the exit of the highest-quality seller leads to a reduction in the market price. The seller's exit triggers another exit wave by sellers with somewhat lower-quality assets, resulting in a further decline in the market price. Investors who are not LPs in a limited partnership fund are faced with the lemon problem, because they have less information about the quality of the fund. This lack of information makes them less willing to pay what might actually be a fair price for a fund share, and under certain conditions they might not be willing to pay for a share at all. Some specialized market participants maintain proprietary databases with detailed intelligence on funds and proactively seek to source deals. In some cases, these buyers may even possess superior information about a fund or a portfolio of funds an LP wants to sell.

13.2 PRIVATE EQUITY CASH FLOW SCHEDULES

Though it may sound simple for an investor to make a 10% portfolio allocation to private equity, the mechanics of the asset class render this level of precision impossible. Investors have little control over the timing of their PE investments and the return of capital from those investments. Access to specific managers and investment styles is not always possible, as managers periodically raise assets for funds in future vintage years, such as every third year, while the market environment may make it easier or more difficult for managers to raise capital for investment in a specific stage, sector, or geography.

13.2.1 Investment Period

A cash flow schedule for a single investment may evolve in the following manner: The general partner embarks on fundraising conversations in January 2012. The fund solicits preliminary commitments during the entire year, requesting that investors sign commitments by December 31, 2012. The fund starts operating the following year and is labeled a 2013 vintage year fund with a stated three-year investment period. A given investor signs a $10 million commitment to the fund. The general partner agrees to draw down capital, which means to request cash from a limited partner only upon identification of suitable investment opportunities or to cover management fees or expenses. As a result, the manager draws down $2 million in 2013, $3 million in 2014, $2 million in 2015, and up to an additional $3 million before the fund's end of life. The size and the timing of these drawdowns are determined exclusively by the general partner, as long as the amount does not exceed the size of the limited partner's commitment. Available evidence shows that for U.S. and European venture capital (VC) funds, the majority of the capital is called between the first year and fifth year of the fund's life. It must be noted that there is substantial variation in the speed of capital calls over vintage years and geographies.

It is important to note that private equity portfolio companies are generally recipients of investor capital in distinct stages. **Follow-on funding** occurs when a particular fund makes an investment subsequent to having been a participant during a prior round of financing. This source of funding is beneficial to both partners. The LP saves on due diligence and is able to allocate capital to a GP who is already familiar to the LP. The GP will benefit as well, as access to other sources of funding may take time.

13.2.2 Harvesting Period

Exits are even more difficult to predict, as both the timing and the size of an exit are highly uncertain. A VC manager takes an initial stake in the company during the investment period of the fund, perhaps in the third year. Ideally, the fund holds the investment for a five-year period, in which company growth, innovation, operational improvements, and financial engineering add value to the underlying investment. The **harvesting period** of an investment occurs in the later years of the fund, after the investment has matured and changed in value, and marks the time when the fund seeks to exit its investment. The value of the investment is related to the growth of the company, the value added by the fund manager, and the market environment at the time of the exit. Empirical results tend to show that exits are quite variable, with the majority of distributions from venture funds occurring after year 6.

In order for an investor to maintain an average allocation of 10% to private equity, it will be necessary to estimate the speed and size of drawdowns as well as the size and timing of exits. Given that vintage years and strategies have quite different experiences, investors need to diversify their private equity allocations over several vintage years and be prepared to have an allocation range, rather than an exact target, for the size of their allocation to private equity.

13.3 FIVE SOURCES OF LIQUIDITY

Commitments are generally met through cash inflows, supplemented by assets readily convertible to cash, or through a company's capacity to borrow. To achieve a competitive total return on committed capital, the investor needs to manage both the investment of uncalled capital during the drawdown period and the reinvestment of distributed capital. The maturity structure of treasury assets (i.e., investments that will fund PE commitments) and the timing of needed private equity funds should ideally be matched, and there should be well-diversified and stable sources of funding, such as the following:

1. Liquidity lines. A short- and medium-term borrowing facility can be managed by the limited partner. Cash needs to be available to meet capital calls, but should cash resources run out, a **liquidity line**—a line of credit secured from external sources—is used. Structuring a sensible liquidity line needs to reflect such factors as the expected amount and timing of cash needs, loan terms, leverage, other services offered, and possibly the rating of the liquidity provider.
2. Maturing treasury investments. While it is tempting to maintain undrawn capital in short-term instruments, such a policy is likely to adversely impact total return. Treasury investments, which are allocations of uncommitted capital to investment products, can be used as a source of liquidity. To achieve higher returns, the expected cash flow demands (liabilities) of the private equity fund should be modeled and matched with those of a treasury portfolio that is composed of assets with the same maturities but that is also expected to produce a return exceeding short-term instruments. Because of the inherent uncertainties related to timing and amount, a given maturity structure of treasury assets can only serve as an approximate match with a fund's capital calls.
3. Realizations of other investments. With the potential for mismatch with cash flows received from treasury assets, the risk of illiquidity is present. Therefore, plans to fund drawdowns may occasionally require a liquidation or realization of other existing positions. To limit the potential for losses resulting from market fluctuations, strict criteria for the eligibility of investments should be applied. Typically, assets that are liquid and relatively stable are used for this purpose. Prices of such assets tend to rise during periods of financial distress.
4. Sell-off of limited partnership shares. Private equity funds are illiquid investments, and early redemption is usually not allowed. Limited partners are generally prohibited from transferring, assigning, pledging, or otherwise disposing of their limited partnership interests or withdrawing from the partnership without the prior consent of the general partners, who can grant or withhold consent at their sole discretion. However, there is a growing secondary market where seasoned fund investments may be liquidated. Therefore, **sell-off of limited partnership shares**, in which sale of shares in the partnership takes place in secondary markets, can be a source of liquidity. But a realization of limited partnership shares in an attempt to increase liquidity is problematic, as it takes considerable time to identify buyers and negotiate the transaction, which may take place at a different price than the NAV of the fund.
5. Distributions from private equity funds. A reinvestment plan should be established that takes into account the uncertainty inherent in the timing and magnitude of distributions from private equity funds. As investments in PE funds are speculative and require a long-term commitment, there is no certainty regarding timing and amounts of distributions. It is also possible that part of the return is received as a distribution in kind in the form of marketable restricted securities. Consequently, reinvestment planning also exposes the investor to considerable liquidity risk.

If several capital calls cannot be met simultaneously, the last resort is for the investor to default. However, in addition to the damage to reputation suffered by the defaulting limited partner, there are stiff penalties associated with not meeting a drawdown request. These penalties include the termination of the limited partner's right to participate in the fund's future investments, the loss of entitlement to distributions or income, the mandatory transfer or sale of the LP's partnership interests, the continuing liability for interest on the defaulted amount, the partial or total forfeiture of the partnership interest, and the liability for any other rights and legal remedies the fund managers may have against the defaulting investor. Defaulting limited partners may continue to be liable for losses or expenses incurred by the fund.

13.4 INVESTMENT STRATEGIES FOR UNDRAWN CAPITAL

Program managers should only be given resources that are necessary for investing in private equity funds. If large liquidity buffers are planned, the program's return target requires that undrawn capital be managed to maintain liquidity while reducing the opportunity costs of investments in lower-return assets.

To minimize the amount of idle capital in the hands of the fund, drawdowns should ideally be just-in-time or possibly also subscription based, and an overcommitment strategy should be followed (see section 13.7). During situations in which large positions of undrawn capital cannot be avoided, the return could be increased by maximizing exposure either to long-term bonds, which requires predictability and planning, or to other higher-yielding asset classes. For this purpose, some authors have suggested public equity (e.g., see Kogelman 1999 and Oberli 2015).

The management of undrawn capital should be left to the limited partner, who will determine if flexible follow-on financing can be arranged. In any case, the premium between liquid assets and illiquid private equity funds sets relatively low upper bounds for returns on the undrawn capital. Alternatively, capital can be put to work more efficiently by investing in other assets with respectable return expectations if they fall into the fund's core expertise, such as publicly quoted private equity or other liquid alternative assets.

13.5 MODELING CASH FLOW PROJECTIONS

Investments in illiquid assets present particular challenges for portfolio management, as there is a high degree of uncertainty inherent in the timing and amount of cash flows.

13.5.1 Long-Term Management of Investment Commitments

Strategic commitment steering is the long-term management of investment commitments, with the main objective of building up and maintaining a balanced and stable portfolio in line with the investment strategy. The portfolio balance and allocation depend not only on the level of undrawn commitments to fund investments but also on the rate and timing of drawdowns and distributions. Effective management of a private equity program requires a reasonably accurate assessment of the individual fund's future cash flow pattern to enable the steering of commitments and portfolio balance. Maximizing the return on undrawn commitments often requires taking positions in assets with limited liquidity. This process needs to be started as early as possible, because a profitable realization of such positions may take some time. Thus, the establishment of efficient treasury management also relies heavily on projection and planning methodologies.

Projection models should not be needlessly complicated. They should be sensible on a theoretical basis, and should be able to incorporate and respond to actual cash flow experience and valuations. Such models should also be able to analyze the portfolio impact of varying return scenarios and changing rates of investments and repayments. Projections need to consider existing deals with known characteristics as well as future deals with unknown characteristics or characteristics yet to be chosen (such as commitment levels).

13.5.2 Four Inputs to Projection Models

Sophisticated alternative investment and fund-of-funds managers have developed four proprietary approaches[4] that take a series of inputs into consideration:

1. Market and empirical data. These data come mainly from data services such as Bloomberg, but internal data can also form the statistical inputs for forecasting expected drawdowns and distributions. These are complemented by an assessment of the vintage-year quality and the investment environment, along with empirical data to model expected drawdowns, repayments, and exit scenarios.
2. Expert judgment. Significant judgment is required for estimating inputs for determining valuations when applying quantitative approaches. A high level of knowledgeable expertise gained through industry experience is beneficial in accurately interpreting empirical and observed information, especially within an opaque environment such as private equity, where often limited data points are available.
3. Fund data. Data on actual drawdowns and distributions form the basis for the valuation of individual private equity fund investments (monitoring input on actual drawdowns and distributions). Also, monitoring of the quality of the private equity fund portfolio is of tangential relevance; for example, although write-offs are not immediately relevant for cash flows, they can reduce further financing needs.
4. Models. Projections are generated with the help of various models. Generally, the accuracy of predictions is higher for mature programs than it is for programs in their initial years, a fact attributable to better and more complete data being available, combined with short-term luck playing a lesser role with the passage of time.

13.6 THREE APPROACHES TO FORMING MODEL PROJECTIONS

Very simplistically, one can differentiate three approaches to carrying out projections:

1. Estimates use an assessment of current conditions to identify possible future events. The priority is accuracy, which implies a relatively short time horizon and an intensive data-collection process.

EXHIBIT 13.2 Approaches to Projecting Cash Flows

	Tactical Estimates	Forecasts	Strategic Scenarios
Term	Short-term (3–6 months)	Medium-term (1–2 years)	Long-term (over 2 years)
Based on	Current market situation	Specific market environment	Uncertain market environment
Approach	Data gathering and analysis	Quantitative modeling	Planning

2. Forecasts go beyond the short-term horizon, primarily relying on trend-based analysis. Often expert opinion or crowdsourcing is preferred when making an assessment concerning the continuity or modification of current trends.
3. Scenarios can be thought of as a range of forecasts, but both their construction and their intent are more complex. They aim to describe different environments based on plausible changes in current trends.

There is a significant difference between scenarios and forecasting. Presumably, forecasts are attempts to predict the future, whereas scenarios aim to enable better decisions about the future. Of course, the distinction is not as clear-cut, as depicted in Exhibit 13.2, and typically projection tools combine elements of all these approaches.

13.6.1 Basing Model Projections on Estimates

To estimate is to form an opinion based on imperfect data, comparisons, or experience. Because statistics are of lesser value over the short term (or for a single position), estimation techniques can be more meaningful than forecasts. Estimates can be applied to new commitments in private equity funds to be signed within the next few months and to liquidity events in the near future within private equity funds already committed, as follows:

- New commitments in private equity funds and their first drawdowns can be derived from deal pipeline analysis, with reasonable accuracy for a period of three to six months. Investment managers are typically already in discussions with potential investors. They have a good understanding of the current fundraising environment, the resulting likelihood of commitments, and the size of these commitments.
- There is a series of liquidity events that is either known or reasonably expected to happen. Occasionally, exits (e.g., in the form of initials public offerings [IPOs]) are publicly announced, and possible price ranges are discussed.
- A regularly updated calendar of such events forms the starting point for estimating short- and medium-term liquidity needs. Such an approach is also appropriate for pricing secondary transactions, but in this case, estimates are used until full liquidation of the positions. The use of this technique for long-term purposes tends to be quite time-consuming and information-intensive, leading to significant costs that cannot necessarily be justified for classical portfolio projections; however, it may be appropriate in the case of secondary transactions, in which higher benefits can be expected.

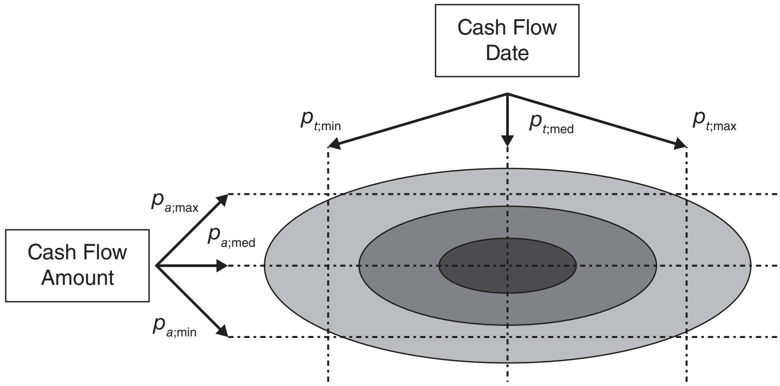


EXHIBIT 13.3 Estimate Grid of Cash Flows, Dates, and Probabilities
Source: Mathonet and Meyer (2007).

13.6.1.1 An Example of Estimation Techniques Distributions from private equity funds to their investors are more sensitive to short-term information and changes. Estimates can be significantly improved through closer interaction with general partners and by incorporating judgment.

A number of firms have developed cash flow models with the objective of generating the most accurate projections of net capital flows from managed funds to investors as possible.[5] In order to measure the uncertainty of exit values and dates, the models incorporate a probabilistic methodology. In such models, investment managers are routinely asked to provide earlier, median, and latest exit dates, and minimum, median, and maximum exit values, as well as the attached probabilities for each event and the basis for the estimates. Such an estimate will take the form shown in Exhibit 13.3, with:

$$P_{a;\min} + P_{a;\text{med}} + P_{a;\max} = 1 \quad \text{and} \quad P_{t;\min} + P_{t;\text{med}} + P_{t;\max} \leq 1 \qquad (13.1)$$

where $p_{a;\min}$, $p_{a;\text{med}}$, and $p_{a;\max}$ are the probabilities of the minimum, median, and maximum cash flow amounts occurring, and $p_{t;\min}$, $p_{t;\text{med}}$, and $p_{t;\max}$ are the probabilities of the earlier, median, and latest cash flow dates occurring. This assumes that the cash flows can be approximated by three buckets.

The fact that the probabilities for cash flow dates do not necessarily add up to 1 allows for the situation in which the cash flow is not certain to take place at all, in which case no probability would be assigned to the latest date.

This approach calls for the following comments:

- The accuracy of the model's projections depends on estimates for price and date of portfolio exits and, therefore, is subject to the uncertainties of market conditions and buyer sentiment.
- Estimates of exit values and dates are uncertain variables. Since values are based on established valuation methods, estimates of exit values tend to be more accurate than estimates of exit dates. For example, a multiple of certain accounting metrics might give a relatively accurate estimate of the value, but there is no established model to estimate the timing of the exit.

- Exit dates should be determined using estimation methods covering only a short time frame. Typically, after more than one year, estimates cease to be of relevance, and forecasting techniques become more important.

A limited partner needs to overcome significant barriers to access timely information of sufficient quality to conduct a bottom-up analysis that will produce precise projections. Due to the lack of detailed information, however, a bottom-up analysis is usually possible only with limitations, often resulting in estimates lacking accuracy.

While fund managers will feel reluctant to disclose information on likely financing or exit events before deals actually close in order to protect their negotiation positions, some general terms are made public in order to secure external funding. Consequently, this is not made part of the standard reporting, and only in rare cases, such as with an IPO, are investors informed in advance. This does not preclude rumors and leaks that may take place in the media. Also, there are various analytical techniques that can be used to form educated guesses. The basis for such an estimate could be discussions with the fund's management on possible new investments or planned realizations, or an assessment of the maturity of the fund and the current exit conditions.

The number of estimates can become quite significant, even for a small portfolio. The combination of three exit values and three dates yields nine possible outcomes per private equity portfolio company. Combining only 10 such companies, with each having three possible input estimates (minimum, median, maximum) for both exit values and dates, into one private equity fund would result in $(3 \times 3)^{10} = 3.5$ billion different potential exit outcome combinations to analyze.

When aggregating the estimates into projections, the private equity fund's structure is also relevant. For example, the preferred return (or hurdle rate) can heavily distort the cash flow to the limited partner. With 100% catch-up, once the limited partners have the preferred return they cease to share further new realizations until the general partner has made a full recovery.

13.6.1.2 Monte Carlo Simulation and Estimates Monte Carlo simulation is a method for exploring the sensitivity of a complex system by varying parameters within statistical constraints. These systems can include financial, physical, and mathematical models that are simulated in a loop, with statistical uncertainty between simulations. The results from the simulation are analyzed to determine the characteristics of the system.[6]

A relatively simple Monte Carlo simulation greatly helps to determine the most likely probability distribution for our sample's 3.5 billion exit outcomes. With a sensitivity analysis (taking into account longer or shorter time periods and lower or higher realization levels), the robustness of results can be checked using the following three methods:

1. The estimates can be interpreted as inputs for a discrete distribution. Volatility, as expressed by the standard deviation of the returns and, therefore, a judgment on risk, is implied by the inputs.
2. Another approach would be to take the investment managers' inputs as a continuous distribution. Their average estimate is taken, and a normal distribution is

assumed as to both value and timing, with a set standard deviation for both the exit value and the exit timing. The parameters used could be based on historical and contemporaneous evidence for similar funds.

3. A third alternative used is a non-normal probability distribution created on an ad hoc basis, following the shape of the curves defined by the three exit values and/or times and the associated probabilities. Here, each deal can have a different curve.

If applied consistently, this, too, despite its simplicity, can be highly useful in assessing likely cash flow patterns over a short time frame.

13.6.1.3 Implementation Issues Using Estimates For short-term estimates in a bottom-up approach, private equity funds and their portfolio companies are analyzed in detail. A thorough bottom-up analysis is a resource-intensive exercise; therefore, especially for large portfolios, all positions and scenarios cannot be continuously reviewed. Also, different general partners will provide information with varying levels of detail and reliability, with some even ignoring requests. However, in most cases, it is possible to split the portfolio into parts based on higher and lower probabilities of cash flows. As one example, since mature companies in booming market segments are more likely to exit than recently funded and young companies, investors could consider focusing their exit analysis on the parts of the portfolio with higher activity levels. Even in more active market situations, many funds can be eliminated as candidates for detailed analysis up front:

- Funds that are so early in their lives that no positive cash flows can be expected within the foreseeable future
- Funds in later stages but for which the majority of portfolio companies are too young to confidently estimate exit details
- Funds in market segments that are comparatively inactive

As a starting point to narrow the population down to a meaningful list of funds to be analyzed, such a bottom-up analysis should be combined with a macro view that takes into account the different private equity funds' investment styles, geographical orientations, ages, industries, and stages. For the relatively idle part of the portfolio, simplistic techniques like "next quarter's forecast is equal to last quarter's realized cash flow," in combination with medium-term forecasts, may be applied.

Even if estimation techniques occasionally lack precision, they are an indispensable tool for anticipating liquidity shortfalls and serve as an early-warning system to supervise limits. While all investments have to be monitored closely, for an illiquid asset class such as private equity, it is especially critical to continuously monitor developments and initiate changes as early as possible.

13.6.2 Basing Model Projections on Forecasts

Generally, many forecasts are based on the assumption that the past can be extended into the future and built on statistical extrapolation of variables. Such approaches are mainly quantitative and aim to predict over the medium term. For private equity funds, a forecasting approach should include, among other items, such criteria as the fund's investment style, life cycle characteristics, age, empirical data for comparable funds, and market data (e.g., stock market indices).

A private equity fund's life cycle characteristics are typically modeled through cash flow estimates based on historical fund data. This data could be from prior investment experience or provided by data services such as Thomson Reuters, Preqin, Cambridge Associates, or Bloomberg. The underlying assumption is that the pattern (timing, amount of cash flows) is the same for each fund or quality of fund, and that scaling can be applied. This allows for the adjustment of the cash flows for differences in fund size. For example, linear extrapolation would indicate how cash flow would change in proportion to change in the size of a fund. Drawdowns tend to follow a reasonably predictable schedule but show marked differences between investment environments. Limitations, such as the incompleteness of the cash flow library and the unavailability of data in general, pose restrictions. One alternative is to use public market indices as part of an econometric model. Small-company equity markets generally serve as the primary exit vehicle for private equity investments either through an IPO or via company valuation in the case of mergers and acquisitions. To be of any predictive value, econometric forecasts need to consider several explanatory variables; thus, they carry significant model risk, which is made worse by the low availability and quality of private equity data. Therefore, continuous review and validation of input parameters and results are critical. Consequently, it is important that all trend-based forecasting approaches be used only with a stern warning and only in conjunction with alternative approaches, such as reality checks. For example, results should be consistent with the assumed scenario. Blindly following econometric models representing a specific market environment is an accident waiting to happen.

13.6.3 Basing Model Projections on Scenarios

While forecasting gives the most likely picture of the future against which plans can be judged, long-term projections, in particular, are fraught with considerable uncertainty, regardless of the kind of forecasting problem. If the environment changes, statistical extrapolation techniques will likely fail. As forecasts do not communicate uncertainty (especially in the VC industry, which thrives on innovation), reliable forecasting has its natural limits. The nearer term one looks, the more predictable the future. In the short term, most people are inclined to estimate or forecast, whereas long-term planning relies on scenarios about the state of the economy. Scenarios are a set of reasonably plausible but structurally different futures that are a useful tool for setting out a course in the face of uncertainty.

Scenarios can be an individual's isolated opinion or can be discussed in groups. This approach is built on the assumption that some people can be more expert than others in predicting what will happen or in excluding what will not happen. This expertise is based on experience, access to privileged information, or a unique ability to assimilate information from various sources. Long-term scenarios are often abstract and simplified, and do not consider the same level of detail as forecasts. Consequently, their results lack comprehensiveness and often cannot differentiate among the natures of cash flows. The assumptions underlying the scenarios need to be documented. Also, intervention points need to be defined; for example, under which circumstances should a scenario be seen as invalid and a new round of scenario development be required? Scenario development helps users think through the process, better understand the environment, and enhance their capability to recognize unexpected events.

The Yale model was developed by CIO David Swensen to aid in the Yale University endowment's strategic commitment steering (Takahashi and Alexander 2002). Essentially, it is a scenario-based tool that considers NAVs, commitments, drawdowns, repayments, fund lifetimes, exit patterns, and growth rates. It does not give variances but only averages, and the variety of possible outcomes is not described by volatilities but rather approximated through the choice of scenarios.

13.7 OVERCOMMITMENT

As explained in Chapter 8, the cash flow J-curve (see Exhibit 8.4) results from the fact that in standard private equity fund structures, commitments (which are firm) are drawn down as needed (i.e., on a just-in-time basis). Realizations are distributed as soon as is practical. During the life of a fund, the net cash outflow (as well as the market exposure, often measured by the NAV) will not be as high as the initial commitment. The timing and size of cash flows to and from a single private equity fund are not known until they are announced. But for a diversified private equity fund investment program, they follow a predictable pattern that can be reliably estimated, forecasted, or planned.

Ideally, investors should find ways to minimize the opportunity cost of these unproductive resources allocated to private equity. On one side, the time lag between commitment and actual investment can be reflected in the maturity profile of the treasury investments to give some extra basis points of return. On the other side, limited partners can implement an overcommitment strategy in which more than the available resources are committed in order to achieve the target investment or exposure level.

An overcommitment strategy is not simply setting a maximum allocation of commitments per year; it requires being underpinned by a detailed understanding of the cash flow profiles of private equity funds. An **overcommitment ratio** is the ratio of total commitments to resources available for commitment for a private equity investor. Overcommitment ratios of less than 100% suggest an inefficient use of resources; 125% to 140% ratios have been documented.[7]

$$\text{Overcommitment Ratio} = \frac{\text{Total Commitments}}{\text{Resources Available for Commitments}} \qquad (13.2)$$

For example, an investor with a new \$200 million allocation to private equity may decide to commit to \$280 million of new private equity investments in order to be fully invested, given the potential for time lags and undrawn capital. The overcommitment ratio would be 1.40, or 140%, found as \$280 million/\$200 million.

APPLICATION 13.7

An investor seeking to invest \$150 million in private equity commits to investments that exceed the \$150 million by \$50 million. What is the investor's overcommitment ratio?

The overcommitment ratio would be based on having the total committed (\$200 million) in the numerator, not just the excess, using Equation 13.2:

$$\text{Overcommitment Ratio} = \$200\text{ million}/\$150\text{ million} = 1.33,\ \text{or } 133\%$$

The overcommitment ratio can be determined on the basis of empirical data. Assuming that, on average, not more than 70% of commitments are actually called, an overcommitment level of around 140% would be feasible.[8] The hypothetical investor with an overcommitment ratio of 140% who was maintaining a strategic asset allocation of 10% to private equity would sign commitments for private equity investments for 14% of the fund (140% × 10% = 14%). But the reality can be significantly different, and large fluctuations around averages can be observed, notably during some overheated market conditions.

As stated previously, the implementation of a successful overcommitment and investment strategy for undrawn capital largely depends on the quality and precision of cash flow projections. Other sources of liquidity should be taken into account as well. A limited partner's overall portfolio composition has an impact on such projections. In a diversified private equity portfolio, cash flow patterns of various fund types can be exploited within a portfolio approach. Buyout and mezzanine funds typically draw down commitments more quickly than do VC funds, which tend to stretch drawdowns in parallel with their stage financing approach. Buyout and mezzanine funds also tend to start distributing more quickly, as they usually have annual income components, such as interest on subordinated debt or dividends on preferred stock. They also invest in established companies, often resulting in a shorter period to reach exit.

To achieve a higher overcommitment level, the limited partner's portfolio should be diversified over several vintage years. As distributions begin to occur, they theoretically supplement resources available to be spent on new capital calls. If there is no vintage-year diversification in the extreme, all private equity funds achieve their maximum investment level simultaneously. Exhibit 13.4 shows simulated data about the percentage of the commitment that is allocated in each year. We can see that if the portfolio consists of only one vintage year, the percentage allocation will fluctuate more than if the portfolio consists of two vintage years. Further, the diversification benefits are greater if the portfolio consists of two vintage years that are far apart (e.g., a portfolio consisting of vintage years 1 and 5). The lesson is that diversification across vintage years is essential for an overcommitment strategy.

As commitments are firm and private equity is a long-term illiquid asset, an overly aggressive overcommitment and treasury strategy may leave limited partners short of liquidity and expose them to the risk of defaulting on a capital call.

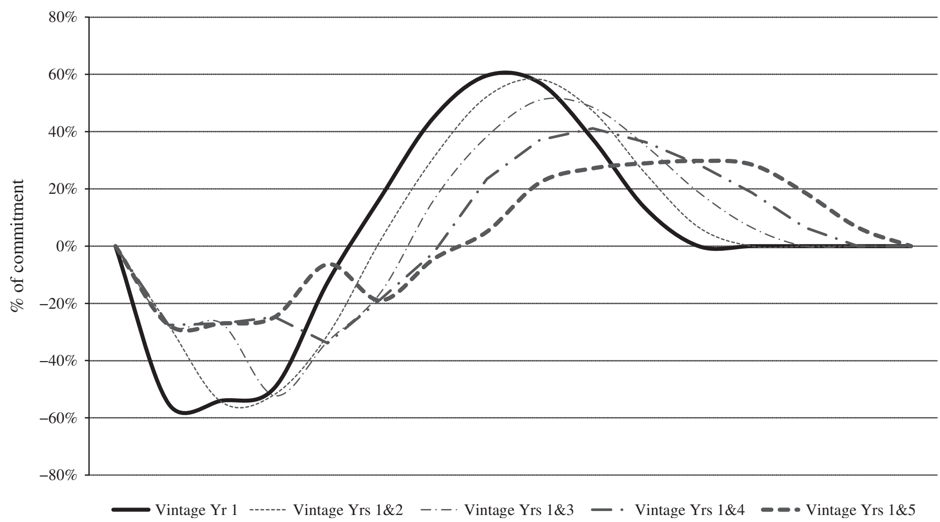


EXHIBIT 13.4 Time Diversification: Cumulative Cash Flows (as Percentage of Commitment) over Time for a Portfolio of Two Private Equity Funds with Commitments Made with a 4-, 3-, 2-, 1-, or 0-Year Gap between Funds
Source: Authors' calculations.

Consequently, especially when sufficient backup facilities are unavailable, having a high percentage of available and allocated resources actually invested in private equity results in a very challenging task, as all these interdependencies will need to be taken into consideration.[9]

13.8 CONCLUSION

To conclude, we would like to make the final observation that, as is often stated in this industry, investing in private equity has a lot in common with winemaking. In talking about wine, some authors argue that only those civilizations that drank wine survived when overall water quality turned poor, as it regularly did. Similarly, we believe that when traditional asset classes enter turbulent market conditions, investors who have invested in private equity (venture capital in particular) and who have used the approaches and tools discussed in this material will be better equipped to navigate through these turbulences.

In private equity, some claim to have found ways to rapidly generate above-market returns, but we remain skeptical. As with making wine, a series of obstacles must be overcome, and if one gives up, the effort may go unrewarded. Inevitably, some vintage years will be disastrous. But a well-stocked wine cellar may help us survive these downturns, and eventually the spectacular vintages may make the effort worthwhile. For these reasons, one cannot switch into and out of winemaking; it is a decision for life (and, most likely, the lives of one's children). The same applies to private equity: To some extent, it can be seen as a decision for one's professional life and, most likely, the life of one's successor.

NOTES

1. We acknowledge the coauthorship of Thomas Meyer and Pierre-Yves Mathonet, who are grateful to Juan Delgado-Moreira, PhD, CFA; Hamilton Lane; and Dr. Michael Jean Gschrei of Dr. Gschrei & Associates GmbH for their valuable comments and suggestions. The views expressed in this chapter are the coauthors' and are not necessarily shared by the aforementioned contributors.
2. See British Private Equity and Venture Capital Association (2015).
3. See ibid.
4. The Partners Group's model (see Wietlisbach 2002) differentiates between strategic and tactical commitment steering. Its private equity management approach rests on the following four pillars: empirical data, actual data, investment advisory, and quantitative management.
5. See Mathonet and Meyer (2007). Similar models are provided by Baring Private Equity Partners, Capital Dynamics, eFront, and others.
6. Wolfram MathWorld.
7. Examples: Schroder Private Equity Fund of Funds (listed on the Dublin stock exchange), 130% overcommitment ratio; VCH Best-of-VC, 140% overcommitment ratio. Hewitt Investment Group recommends overcommitment ratios of 125% to 135% (see Schwartzman 2002).
8. See Schaechterle (2000): "Analysis of statistical data from Venture Economics funds that invest in private equity have a peak level of 65%, leaving the remainder of the investors' committed capital invested in short-term investments. The resulting opportunity costs dilute the overall performance of an investor's private equity allocation by approximately one-third."
9. A real-life case study in Mathonet and Meyer (2007, chap. 4) further highlights the perils of overcommitting and should serve as a warning that relying solely on simple ratios is an insufficient plan for controlling risk.

REFERENCES

British Private Equity and Venture Capital Association. 2015. "Risk in Private Equity." BVCA Research Paper, October.

Kogelman, S. 1999. "The Importance of Asset Allocation in Managing Private Equity Commitments." Goldman Sachs Investment Research.

Mathonet, P., and T. Meyer. 2007. *J-Curve Exposure: Managing a Portfolio of Venture Capital and Private Equity Funds*. Hoboken, NJ: John Wiley & Sons.

Oberli, A. 2015. "Private Equity Asset Allocation: How to Recommit?" *Journal of Private Equity* 18 (2): 9–22.

Schaechterle, S. 2000. "Taking Away the Disadvantages." Partners Group, Baar-Zug, Switzerland.

Schwartzman, T. 2002. "Alternative and Liquid Alternative Assets: Structuring and Oversight." Presentation to Investment Institute's Endowment and Foundation Forum, Atlanta, GA, Hewitt Investment Group, January.

Takahashi, D., and A. Alexander. 2002. "Illiquid Alternative Asset Fund Modeling." *Journal of Portfolio Management* 28 (2): 90–100.

Wietlisbach, U. 2002. "Private Equity Fund-of-Funds Management: A Strategic Approach." AltAssets, London, June.

World Economic Forum. 2011. "The Future of Long-Term Investing." World Economic Forum.

PART 3

Real Assets

CHAPTER 14

Real Estate as an Investment

Real estate has been a very large and important portion of wealth for thousands of years. Even as recently as a century ago, real estate dominated institutional portfolios and was classified as property. During recent decades, the preeminence of real estate has yielded to the growing importance of intangible assets. Yet real estate remains a valuable part of any well-diversified portfolio.

The transition of real estate from dominating traditional institutional-quality investments to being an alternative investment raises important issues in terms of how to evaluate real estate on a forward-looking basis. This chapter provides an overview of the attributes, asset allocation, categories, and return drivers of real estate. The chapter concludes with a discussion of the four-quadrant model, which is a graphical representation of a real estate system.

14.1 ATTRIBUTES OF REAL ESTATE

Real estate—and any other asset, for that matter—should be included in a portfolio until the marginal benefits of additional investment equal the marginal costs of additional investment. An optimized portfolio is achieved when additional investments in each asset class are equally attractive. In other words, exposure to each type of real estate investment, and to real estate overall, should be added until the net benefits have diminished to the point that allocations to other investments are equally attractive.

14.1.1 Five Potential Advantages of Real Estate

What are the aspects of real estate that make it attractive or unattractive relative to other asset classes? There are five common attributes that encourage the inclusion of real estate in an investment portfolio:

1. Its potential to offer absolute returns
2. Its potential to hedge against unexpected inflation
3. Its potential to provide diversification against stocks and bonds
4. Its potential to provide steady cash inflows
5. Its potential to provide income tax advantages

These potential advantages, the first three of which are related to portfolio risk, do not necessarily come without costs. In particular, to the extent that markets are competitive and efficient, market prices of real estate will tend to adjust, such that any relative advantages to real estate will be offset by lower expected returns.

This list of potential advantages to real estate investment is not comprehensive. For example, another motivation would be to own all or part of a trophy property that offers name recognition, prestige, marketing potential, and enhanced reputation to the owner (e.g., a large high-quality office property in a prominent location). Another potential advantage is that real estate typically allows investors to use a high degree of leverage.

14.1.2 Three Potential Disadvantages of Real Estate

There are also aspects of real estate that can discourage its inclusion in an investment portfolio:

1. Its heterogeneity
2. Its lumpiness, which may prevent investors from creating optimal portfolios
3. Its illiquidity, which may reduce the opportunity to rebalance and sell assets at fair market prices in a short period

Real estate is a highly heterogeneous asset. Not only are the physical features of the individual properties unique in terms of location, use, and design, but varying lease structures can lead to large differences in income streams. This heterogeneity is particularly troublesome as it relates to the due diligence process. Accordingly, due diligence of real estate investments can require specialized analysis and managerial skill.

The second potential disadvantage of real estate is lumpiness, including the indivisibility of direct ownership. Lumpiness describes a situation in which assets cannot be easily and inexpensively bought and sold in sizes or quantities that meet the preferences of both buyers and sellers. Listed equities of large companies are not lumpy, because purchases and sales can be easily made in the desired size by altering the number of shares in the transaction. Direct real estate ownership may be difficult to trade in sizes or quantities desired by a market participant. While growth in real estate investment trusts (REITs) and in a number of alternative real estate investment vehicles has led to divisible investment opportunities at an indirect ownership level, investors at the single-property level are still faced with the choice of either buying the entire asset or not buying it. The inherent indivisible nature of individual real estate assets leads to problems with respect to high unit costs (i.e., large investment sizes) and relatively high transaction costs, including those transactions involving joint ventures.

The final major disadvantage relates to the illiquidity of real estate. As a private, non-exchange-traded asset with both high unit costs and high transaction costs, real estate can be highly illiquid, especially when compared to traditional securities. Important implications of illiquidity are its effects on reported returns, extended holding periods, and the ability to transact at reasonable valuations when either demand or supply evaporates.

All three of these characteristics complicate performance measurement and evaluation of real estate investments. The goal of each investor is to find the level and composition of real estate exposure that optimize the portfolio's intended return-risk profile when considering all benefits and costs.

14.2 ASSET ALLOCATION

This section discusses major methods of categorizing real estate and differentiating among real estate investments, the understanding of which leads to two important outcomes. First, a better understanding of the breadth of real estate investment opportunities helps refine an asset allocator's decision as to how much capital to allocate to real estate. Second, a nuanced understanding of the different categories of real estate facilitates the decision of how to allocate funds within the real estate portfolio.

14.2.1 Heterogeneity within Subcategories

Not only is real estate heterogeneous among subcategories, but it can also be highly heterogeneous *within* its subcategories. Although categorization and subcategorization of real estate may serve a useful role in asset allocation and analysis, care must be taken to avoid development of an oversimplified view of real estate. Although assets within various real estate categories and subcategories typically share general characteristics, there may be instances in which tremendous differences in their economic nature exist.

For example, consider two office buildings that are similar in size, construction, and location. The first office building has a 20-year non-cancellable lease with a large well-capitalized and well-hedged corporation. The lease essentially locks in the rental revenues for the entire property for the next two decades. In this case, the annual income of the property will be similar to that of a corporate bond, and the value of the property to the investor will tend to fluctuate in response to the same factors affecting the value of a corporate bond issued by the tenant (i.e., riskless interest rate changes and changes in the credit spread on the debt of the tenant). The principal difference affecting income and valuation between the two is that while many commercial real estate leases allow for periodic rent increases through contractual escalator clauses, coupon payments on outstanding corporate bonds are not inflation-adjusted.

The second office building in the example is vacant. Both buildings are located in a geographic area with an economy strongly linked to oil prices. The value of this empty real estate asset will be especially sensitive to the supply of and demand for office space in the local real estate market. Thus, the value of this property will be driven by the forces that affect the region's economy—in this case, oil prices. The vacant property's value may behave more like equity prices in general and like oil stock prices in particular. However, if the building begins to attract tenants with long-term, non-cancellable leases, the property's fundamental economic nature may transition from being more like an oil stock to being more like a corporate bond. (The value of the first building will become increasingly driven by the difference in local market fundamentals relative to those implied by the current lease terms as the maturity date approaches.)

This example shows that assets within a specific type of real estate (e.g., private commercial real estate) may behave like debt or equity securities depending on the characteristics of the individual properties. Furthermore, a particular property may experience dramatic changes in its investment characteristics due to a specific event, such as the signing or termination of a very long-term, non-cancellable lease.

14.2.2 Top-Down Asset Allocation

Asset allocation approaches differ by the extent to which the process is focused on top-down allocation versus bottom-up allocation. **Top-down asset allocation** emphasizes allocation based on the analysis of the macro environment and risk premiums, and their expected impact on general categories or types of portfolio investments. Exhibit 14.1 illustrates the concept of target asset allocation using one set of potential categories and weights. To the extent that the allocation illustrated in Exhibit 14.1 is based on general portfolio objectives, the asset allocation process is a top-down asset allocation. Section 14.3 provides numerous distinctions that can be used to place real estate into different categories.

For example, an asset allocator would typically be concerned about the return, risk, liquidity, and even tax implications of the overall portfolio. In the case of liquidity, a top-down asset allocator takes into account the concern for liquidity when selecting portfolio weights, along with a preference for a higher return and other perceptions as to how each category serves the overall portfolio objectives. Thus, in considering liquidity, a pure top-down asset allocator would determine the weights for each category based in part on an analysis of the liquidity of each category of real estate and the extent to which illiquidity offers a risk premium relative to the portfolio's needs for liquidity. Therefore, if the asset allocator's investment policy mandates that the illiquid portion of the portfolio should not exceed some particular level, the

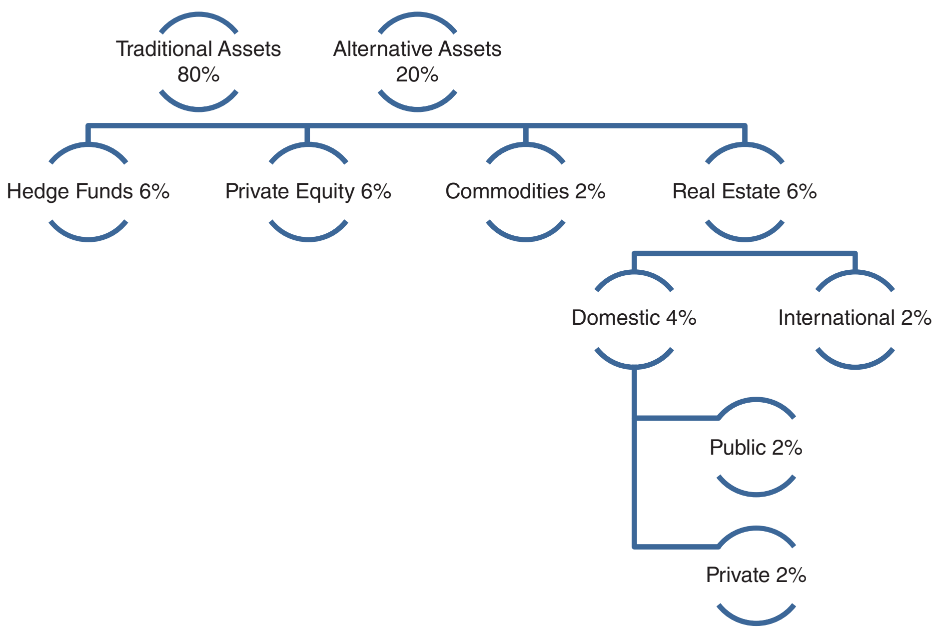


EXHIBIT 14.1 Asset Allocation

asset allocator will have to find the optimal allocation to various categories of real estate subject to this constraint. As a result, this liquidity concern and other general portfolio objectives may cause the asset allocator to make allocations at each level that ultimately determine the feasible and desirable portfolio allocation to domestic, publicly traded equity real estate investment products. It should be noted that further details regarding asset allocation could be illustrated in Exhibit 14.1. For example, in the category of publicly traded domestic real estate equity, the asset allocator may further divide the category by property type (office buildings, industrial centers, data centers, retail, residential, health-care facilities, self-storage facilities, and hotels) and by investment management type (in-house vs. external, active vs. indexed). Furthermore, other categorizations could be used, along with greater refinement of the level of detail (e.g., international vs. domestic could be broken into finer distinctions based on geographical profiles or demographic trends).

14.2.3 Bottom-Up Asset Allocation

Bottom-up asset allocation refers to an emphasis on the relative attractiveness of individual investment opportunities as the primary driving factor of the asset allocation process. The underlying analysis is typically supported by rigorous fundamental analysis. The asset allocator may determine that some subcategories of real estate, particular properties, or publicly traded real estate managers offer exceptionally attractive investment opportunities, whereas other categories, properties, or manager portfolio characteristics are unattractive. To the extent that these analyses of individual opportunities or subcategories exert the dominant effect on the ultimate asset allocation, the asset allocation would reflect a bottom-up strategy.

For example, an asset allocator may have internal staff and established relationships with outside managers that lead the allocator to believe that particular subcategories present the portfolio with attractive opportunities relative to the wider asset class. The asset allocator may favor some opportunities based on extensive experience and knowledge, while avoiding potentially attractive opportunities in subcategories in which the allocator has a limited perceived edge. To the extent that individual asset selection exerts a major effect on the ultimate asset allocations among major categories, the asset allocation process is considered to be bottom-up.

Most asset allocation methods are a mix of top-down and bottom-up, in that allocations among major categories tend to be driven by general portfolio objectives and macro analysis, whereas allocations within subcategories are typically driven by the allocator's perceptions of the opportunities available.

14.3 CATEGORIES OF REAL ESTATE

This section describes the main characteristics of various real estate assets, beginning with four especially common categories that can be used to differentiate real estate:

1. Equity versus debt
2. Domestic versus international
3. Residential versus commercial
4. Private versus public

Each of these categories is briefly discussed in the following four sections, followed by discussions of other methods of categorizing real estate.

14.3.1 Equity versus Debt

The traditional method of distinguishing between equity claims and debt claims is to use the legal distinction between a residual claim and a fixed claim. A mortgage is a debt instrument collateralized by real estate, and real estate debt is typically defined as including all mortgages. Note, however, that mortgages with substantial credit risk can behave more like equity, and equity ownership of properties with very long-term leases can behave like debt. The value of a mortgage is more closely associated with the value of the real estate than the profitability of the borrower.

14.3.2 Domestic versus International

One of the primary motivations to real estate investing is diversification. International investing (i.e., cross-border investing) in general and international real estate investing in particular are regarded as offering substantially improved diversification. However, the heterogeneity of most real estate and the unique nature of many real estate investments make international real estate investing more problematic than international investing in traditional assets. Other challenges include lack of knowledge and experience regarding foreign real estate markets, lack of relationships with foreign real estate managers, time and expense of travel for due diligence, liquidity concerns, political risk (particularly in emerging markets), risk management of foreign currency exposures, and taxation differences. For these reasons, a large share of international real estate investing is done through shares of listed property companies (LPCs), including REITs, in foreign countries. The continuing emergence of derivative products related to real estate investments in particular nations or regions is an important potential opportunity for exploiting the benefits of international diversification without the challenges of direct international investment.

The extent of appropriate international investing depends on the locale of the asset allocator. A UK asset allocator or an asset allocator in another country with a very large economy may be able to achieve moderate levels of diversification without foreign real estate investing. However, an asset allocator in a nation with a small or emerging economy may experience high levels of idiosyncratic risk in the absence of foreign investments.

14.3.3 Residential versus Commercial

One of the most important drivers of the characteristics of a real estate investment is the nature of the real estate assets underlying the investment. A broad distinction, especially in mortgages, is residential real estate versus commercial real estate.

Housing or residential real estate properties: **Residential real estate** or **housing real estate** includes many property types, such as single-family homes, town houses, condominiums, and manufactured housing. The housing or residential real estate sector is traditionally defined as including owner-occupied housing rather than large apartment complexes. According to the Federal Reserve (2015), the aggregate value of all U.S. homes amounted to approximately \$20.7 trillion at the end of 2014, representing an important portion of household wealth. In the United Kingdom, housing values totaled £5.2 trillion as of January 2014, up from £3.6 trillion in 2003 (Savills 2014).

Within residential real estate, institutional investors are primarily concerned with investing in mortgages backed by housing and residential real estate. Ownership in these instruments is usually established through pools of mortgages. The global residential mortgage market had total balances outstanding of $25.7 trillion at the end of 2013 (Market Reports Online 2015). According to data from the Federal Reserve and MarketResearch.com, residential mortgages in major markets accounted for approximately:

- $12 trillion in the United States (end of 2014)
- $2.5 trillion in Japan (2013)
- $1.6 trillion in the United Kingdom (2013)
- $1.4 trillion in Germany (2013)
- $1.2 trillion in France (2013)

with outstanding balance of $7.0 trillion in the rest of the world.

Commercial real estate properties: **Commercial real estate** properties include the following property sectors: office buildings, industrial centers, data centers, retail (malls and shopping centers, also referred to as "strips"), apartments, health-care facilities (medical office buildings and assisted-living centers), self-storage facilities, and hotels. Small properties may be directly and solely owned by a single investor. Alternatively, collections of numerous smaller properties and large commercial properties may be managed by a real estate company, such as a publicly listed REIT, or through private equity real estate funds, which, in turn, are owned by several institutional investors as limited partners. Within commercial real estate, the institutional investor can access opportunities through either debt or equity investments. The volume of transactions fluctuates significantly depending on the stage of the business cycle, but it is generally high enough to support large investments by institutional investors. For example, the commercial real estate investments in 2013 and early 2014 averaged about $100 billion in the United States; $75 billion in Europe, the Middle East, and Africa; and $180 billion in the Asia-Pacific region (Deloitte 2015).

For the most part, residential and commercial real estate require very distinct methods of financial analysis. For example, the credit risk of mortgages on residential real estate is typically analyzed with a focus on the creditworthiness of the borrower. Mortgages on commercial real estate tend to focus on the analysis of the net cash flows from the property.

14.3.4 Private versus Public

Exposure to the real estate market, especially the equity side, can be achieved via private and public ownership. **Private real estate equity** investment involves the direct or indirect acquisition and management of actual physical properties that are not traded on an exchange. **Public real estate investment** entails the buying of shares of real estate investment companies and investing in other indirect exchange-traded forms of real estate (including futures and options on real estate indices and exchange-traded funds linked to real estate).

Private real estate is also known as physical, direct, or non-exchange-traded real estate. Private real estate may take the form of equity through direct ownership of the property or debt via mortgage claims on the property. The private real estate market comprises several segments: housing or residential real estate properties, commercial real estate properties, farmland, and timberland. The relative advantages of investing in the private side of real estate equity are that investors or investment managers have the ability to choose specific properties, exert direct control of their investments, and enjoy the potential for tax-timing benefits.

Public real estate is a financial claim in the form of equity, debt, funds, or derivative positions, and may be a claim on either underlying private real estate positions or underlying public real estate positions. Public real estate is also known as securitized, financial, indirect, or exchange-traded real estate. Thus, public real estate enables the ownership of private real estate through one or more levels of contracts designed to facilitate real estate ownership, reduce costs, or increase liquidity relative to direct ownership. For example, securitization, particularly in the form of commercial mortgage-backed securities (CMBS), has substantially increased the liquidity and accessibility of real estate investments, while allowing structured investments based on specific institutional investor return-risk profile targets.

REITs are securitized pools of real estate that constitute an important form of public real estate, especially in the United States, which has the largest market in the world, followed by Australia, France, the United Kingdom, and Japan. The relative advantages of investing in public commercial real estate (as opposed to private) include liquidity, greater investor access, relatively low transaction costs, the potential for better corporate governance structures, and the transparency brought by required Securities and Exchange Commission (SEC) filings and pricing in public capital markets (Idzorek, Barad, and Meier 2007).

One of the most important characteristics of REITs is that, due to the trust structure, income distributed by a REIT to its shareholders is taxed not at the REIT level but at the investor level after it flows through the REIT. In order to enjoy this tax status in the United States, REITs are subject to two main restrictions: (1) 75% of the income that they receive must be derived from real estate activities, and (2) the REIT is legally obligated to pay out 90% of its taxable income in the form of dividends. Other restrictions relate to the ownership structure of the REIT. As long as a REIT is in compliance with the relevant restrictions, it may deduct dividends from its income when determining its corporate tax liability (i.e., it pays corporate income taxes only on retained taxable income). REITs enjoy similar tax efficiencies in other parts of the world—for example, in the United Kingdom and in Germany.

In the United States, REITs can invest both in the private real estate market (equity REITs) and in real estate–based debt (mortgage REITs), though, practically speaking, each specific REIT tends to manage its portfolio by focusing nearly exclusively on either equity ownership (including the use of joint ventures) or debt (including derivatives). Generally, if a REIT has 50% or more of its assets in the private real estate equity market, it is viewed as an equity REIT; if over 50% of its assets are invested in real estate debt, it is viewed as a mortgage REIT. This distinction is important in return analyses. For example, unlike equity REITs, mortgage REITs tend to move in line with other rate-sensitive securities due to their underlying asset base. Equity REITs dominate the REIT sector in terms of both number available and market capitalization. In the United States, as of the end of 2015, equity REITs, as

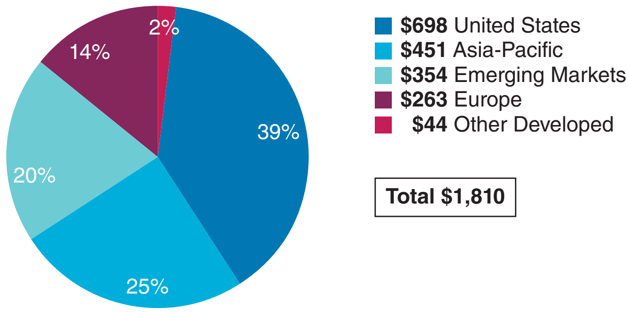


EXHIBIT 14.2 Global Real Estate Securities Market by Region on September 30, 2015; Market Capitalization in US$ billions
Sources: Cohen & Steers, FTSE, FactSet, and Standard & Poor's.

proxied by the FTSE NAREIT All Equity REITs Index, had a market capitalization of $882 billion, whereas mortgage REITs (the FTSE NAREIT Mortgage REITs Index) had a market cap of $52.5 billion (National Association of Real Estate Investment Trusts 2016). And globally, as of September 30, 2015, Cohen & Steers put the total market capitalization for real estate securities at $1.81 trillion across 487 companies (see Exhibit 14.2).

14.3.5 Real Estate Categorization by Market

Institutional investors often categorize private commercial real estate equity investments by the size of the real estate market in which the property is located. Real estate assets are said to trade in the **primary real estate market** if the geographic location of the real estate is in a major metropolitan area of the world, with numerous large real estate properties or a healthy growth rate in real estate projects. Primary real estate markets tend to have easily recognizable names. Using the United States for illustration, examples range from cities such as Orlando, Florida, to specific metropolitan areas, such as Manhattan in New York City. Large institutional investors focus on investments in these primary markets. **Secondary real estate markets** include moderately sized communities as well as suburban areas of primary markets. **Tertiary real estate markets** tend to have less recognizable names, smaller populations, and smaller real estate projects.

14.3.6 Risk and Return Classifications

From the perspective of an asset allocator, the most useful categorization approach of real estate should center on the most important characteristics of the portfolio. The primary characteristics of a portfolio are risk and return. Therefore, asset allocators should consider including a categorization approach that focuses on the risk and return profiles of the assets.

Chapter 16 discusses a classification of real estate that includes three types: core, value added, and opportunistic. These three categories assist the analysis of real estate and asset allocation decisions by grouping together real estate products that possess similar risk and return properties. The heterogeneity of real estate within each subcategory may lead an asset allocator to focus on distinguishing investments based on their risk and return, perhaps using the core, value-added, and opportunistic categories to classify them.

14.3.7 The Focus on Private Commercial Real Estate

Most of the focus in Chapters 15 through 18 is on private commercial (i.e., income-producing) real estate rather than on public real estate, residential real estate, or commercial mortgages. There are three reasons for this:

1. Most commercial real estate throughout the world is privately held rather than publicly traded.
2. Most of the equity of residential real estate is held by the occupier of the property rather than by an institutional investor.
3. The pricing of the equity claims to private commercial real estate drives the pricing of the credit risk in the pricing of commercial mortgages. In other words, real estate debt may be viewed through the structural model as being well explained through an understanding of the risks of the equity in the same property.

Thus, in our remaining four chapters focusing on real estate, the material emphasizes the risk and returns of equity ownership of private commercial properties, whether owned directly or held through limited partnerships.

14.4 RETURN DRIVERS OF REAL ESTATE

Real estate returns are generally perceived as being fundamentally different from the returns of other assets. For example, real estate is generally believed to offer substantial inflation protection and distinct diversification benefits. These distinctions are often justified based on historical tendencies, derived through empirical analysis of past prices and returns.

An understanding of the inflation-protection potential of real estate requires a distinction between anticipated inflation and unanticipated inflation. The **anticipated inflation rate** is the expected rate of change in overall price levels. Expectations vary across market participants and are generally unobservable. Accordingly, indications of anticipated inflation are often based on surveys of consensus estimates, derived from past inflation, or inferred from other market information such as interest rates.

To the extent that a market is informationally efficient, the level of anticipated inflation should already be incorporated in the price and, thus, the expected rate of return on various assets. For example, the **Fisher effect** states that nominal interest rates equal the combination of real interest rates and a premium for anticipated inflation (while other models include the effect of expected taxation):

$$\begin{aligned}\text{Nominal Interest Rate (ex ante)} = {} & \text{Real Interest Rate (ex ante)} \\ & + \text{Anticipated Inflation}\end{aligned}$$

The net result is that every asset in an informationally efficient market provides identical protection from anticipated inflation, since every asset's price adjusts to compensate the buyer for anticipated inflation. Thus, stable or previously anticipated inflation rates should not be a return driver, or determinant, by themselves.

The more challenging issue is that of unanticipated inflation. **Unanticipated inflation** is the realized rate of inflation minus anticipated inflation:

$$\text{Unanticipated Inflation} = \text{Realized Inflation Rate} - \text{Anticipated Inflation}$$

The effect of unanticipated inflation on an investment's realized return is crucial, and the risk of unanticipated inflation is an important consideration in risk analysis. Realized inflation in a particular period exerts its primary effect through its role in modifying future expectations of inflation. Because changes in expected inflation can exert substantial effects on prices, realized inflation can be an important driver of most real estate returns. In other words, deviation in realized inflation rates relative to previously anticipated inflation rates (i.e., unanticipated inflation) can be a very important return driver due to its role in changing anticipations of future inflation rates. Volatility of the inflation rate is another source of risk related to inflation. Finally, unanticipated inflation can have differential effects on relative prices, which represents another source of risk. For example, higher unanticipated inflation may lead to an immediate rise in labor costs while having a muted effect on certain real estate properties.

The sensitivity of various real estate investments to unanticipated inflation may be analyzed through empirical analyses of past returns or a fundamental analysis of the investment's sources of risk and return. A challenge in empirical analysis of the effects of unanticipated inflation is in developing an objective and accurate estimate of the consensus-expected inflation rate. An estimate of the anticipated inflation rate is necessary to estimate the unanticipated inflation rate as the difference between the realized and expected inflation rates.

Another challenge is that there are typically different rates of anticipated inflation over different time horizons. Changes in anticipation of inflation over various time horizons could be expected to exert different effects on various real estate investments. Furthermore, realized rates of inflation may be studied over various time intervals. Thus, the price reaction and inflation protection offered by an investment should be expected to differ based on whether the realized inflation signals a long-term change in expected inflation or a more transient shift in inflation anticipation.

Inflation may have different effects on different types of properties. Inflation may hurt the value of bondlike properties that have long-term leases at fixed rates. Though properties are often valued on a pre-tax basis and before financing costs (interest), investors in real estate equity that are leveraged with adjustable-rate mortgages may also suffer from higher financing costs during times of inflation. Likewise, higher inflation can benefit owners of leveraged properties that are financed with fixed-rate debt, while lower inflation can harm them. Properties with lease structures that may benefit from high inflation include those with short-term leases or leases in which payments contractually rise with the rate of inflation. In fact, many leases governing long-term commercial real estate leases in the United States contain an **escalator clause**, which periodically adjusts lease payments based on some agreed upon measure of inflation.

The following factors also influence real estate returns: the state of the economy, demographics, interest rate level, the tax treatment of real estate income and financing costs, and trends. The next section examines the impact that some of these factors may have on the real estate market, the asset market, and the development (construction) industry in the framework of the four-quadrant model.

14.5 THE FOUR-QUADRANT MODEL

In this section, which draws from Geltner, Miller, Clayton, and Eichholtz (2014), we describe the **four-quadrant model** of DiPasquale and Wheaton (1992), which allows for the simultaneous assessment of the long-run equilibrium within and between the real estate space and asset markets. A **real estate system** consists of three components: the market for real estate space, the asset market, and the construction industry. The four-quadrant model is a graphic representation of the dynamics of a real estate system; hence, it is also referred to as a systems dynamic model.

Exhibit 14.3 shows the four quadrants of the model. The two right-hand quadrants correspond to the property market for the use of space, while the two left-hand quadrants characterize the asset market for the ownership of real estate. The four-quadrant model and the real estate system can be explained as follows:

1. Property market rent: This is the equilibrium rental rate and is determined through the interaction of supply of space and demand for space. It is typically assumed that the supply of space is fixed in the short run. Demand, on the other hand, is a function of economic conditions as well as the characteristics of the property. This appears in the northeast (NE) quadrant.
2. Asset market value: The equilibrium rent of the previous step determines the income that a property generates. Assuming relatively fixed income in the short to medium term, the market value of the property can be calculated as the present

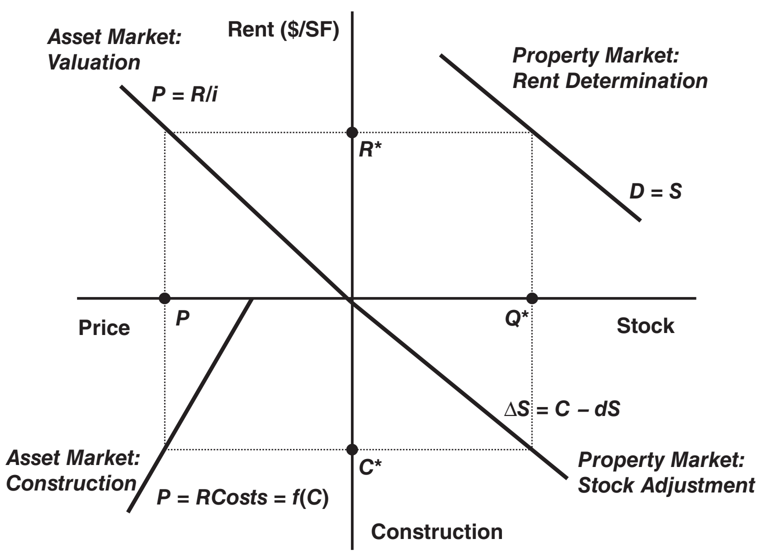


EXHIBIT 14.3 The Four-Quadrant Model
Source: DiPasquale and Wheaton (1992).

value of future cash flows. Given the required capitalization rate of a property, its market value will be equal to net operating income divided by the capitalization rate. Rental income is a major determinant of net operating income, and the current economic condition and the rate of return available on other investments are the major determinants of the capitalization rate required by an investor's other investments. This appears in the northwest (NW) quadrant.

3. Construction activities: Given the equilibrium value of the property determined in the previous item, the construction industry will compare the market value to construction costs. This will be one of the factors that will determine the level of construction activities in the economy. Of course, other variables—such as the availability of credit, economic conditions, and demography—will affect construction activities as well. This appears in the southwest (SW) quadrant.
4. Property market: The construction activities of the previous item determine how the supply will change, eventually affecting the supply space, which was assumed to be fixed in the short run. This means that while the supply is fixed in the short run, the level of rental income will spill over into the construction industry, which with some time lag will affect the supply of space. This appears in the southeast (SE) quadrant.

Now that we have discussed the overall logic of the model, we next provide a more detailed analysis of the four-quadrant model.

The northeast (NE) quadrant has two axes: rent (which is measured in dollars per unit of space, such as square feet) and the stock of space (measured in the same units of space). This quadrant describes how rents in the space market are established. In equilibrium, the demand for space, the downward-sloping line D, is equal to the stock of space, S. Assuming that the stock of space is fixed (a plausible assumption for the short run), rents must be determined so that the space demand is exactly equal to the stock of space. Demand is a function of rent and conditions in the economy. Rent (R^*) is found by plotting a level of stock of space on the x-axis up to the demand line and over to the y-axis. In the short run, rents are determined in the NE quadrant.

The northwest (NW) quadrant has two axes: rent ($) and price ($/SF, or dollars per square foot). The ray starting off out of the origin represents the ratio of rent to price for real estate assets. The ratio of net operating income to price is known as the capitalization rate (i). Here we use the rental income as a proxy for net operating income. Generally, the capitalization rate is taken as exogenous and is a function of the following four variables: the expected growth in rents, the risks related to the income stream from rents, the long-term interest rate in the economy, and the tax code treatment of real estate income. A higher capitalization rate is represented by a clockwise rotation in the ray, and a lower rate by a counterclockwise rotation. The NW quadrant takes the rent level (R^*) from the NE quadrant and establishes a price for real estate assets (P^*) based on the capitalization rate (i). That is, given a required income yield and given the income generated by a property, we can value the property as rent/capitalization rate.

In the four-quadrant model, the price of the real estate asset is determined by moving first from the vertical axis (rent level) in the NE quadrant over to the ray in the NW quadrant, and then down to the price level on the horizontal axis. Continuing moving in a counterclockwise manner, the next quadrant (southwest, or SW) is the section of the asset market where the construction of new real estate assets is determined. Here, the line $f(C)$ represents the replacement costs (*RCosts*) of real estate. Construction costs increase with greater construction activity, and therefore the curve moves in a southwesterly direction, intersecting the price axis at the minimum dollar value per unit of space required to generate construction activity.[1] The level of new construction is established where asset prices equal replacement costs. This is determined by a move to the SW quadrant, where the replacement cost curve determines the level of construction given the price of real estate assets from the NW quadrant. Higher levels of construction would be unprofitable, whereas lower levels would generate excess profits. Therefore, real estate prices (P) equal construction costs (*RCosts*), both of which are functions of the construction level (C).

In the southeast (SE) quadrant, the long-run stock of space is created by the annual flow of new construction. The change in the stock of space in a certain period (ΔS) is equal to new construction (C) minus the depreciation rate (d, which represents losses from the stock) times the stock (S). The ray emanating from the origin corresponds to that level of the stock of space (on the x-axis) that requires a level of annual construction for replacement just equal to that value on the y-axis. At that level, the stock of space will remain constant over time, given that depreciation will be exactly equal to new completions. Therefore, Δ is equal to zero and $C = dS$.

14.5.1 An Illustration of the Four-Quadrant Model: Explaining Real Estate Market Booms and Busts

Let us assume that the real estate market is experiencing a boom, and that demand for space surges unexpectedly (for example, due to an unanticipated increase in employment, production, or the number of households). In Exhibit 14.4, this is represented by a shift to the right in the demand curve for rents in the NE quadrant (from D_0 to

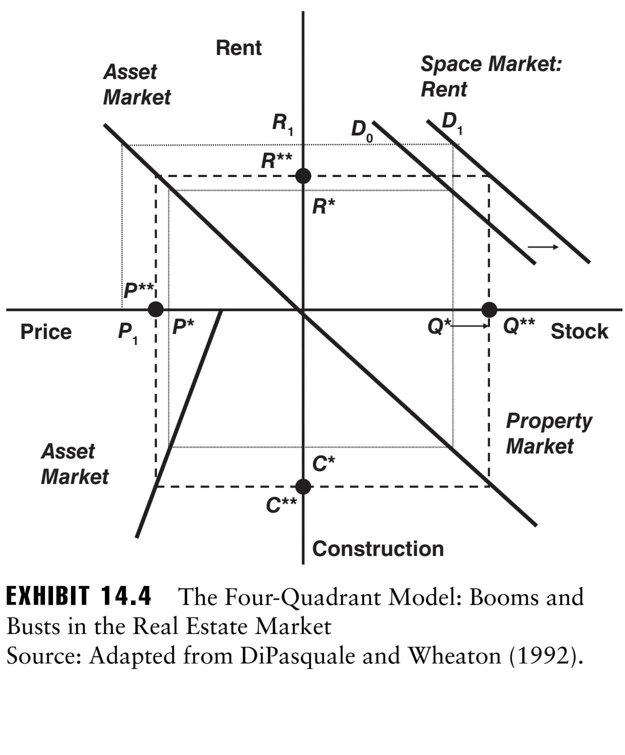


EXHIBIT 14.4 The Four-Quadrant Model: Booms and Busts in the Real Estate Market
Source: Adapted from DiPasquale and Wheaton (1992).

D_1). In the short run, rents rise from R^* to R_1 (the rent level that relates the original stock of space Q^* to the new demand function D_1), and real estate prices increase from P^* to P_1 in the asset market (NW quadrant), as there is no time for new space to be constructed in response to the unexpected increase in demand. However, this is not a long-run equilibrium. After a year or two, new space is developed, and rents decrease from R_1 to R^{**} (and real estate prices from P_1 to P^{**}), which is nonetheless still above the initial equilibrium rent of R^* (and the initial equilibrium real estate prices, P^*). This occurs because the long-run marginal cost function of the real estate market is outward-sloping (SW quadrant). Furthermore, notice that $Q^{**} > Q^*$, and thus the long-run equilibrium quantity of space is higher than it was at the original level of demand for rental space. This result will hold as long as the demand for rents exhibits at least some price elasticity.

We have just observed that, in the short run, an unexpected increase in demand for rental space (NE quadrant) leads to an increase in rents, given that the level of real estate space is fixed. These higher rents then cause a rise in real estate prices (NW quadrant), which, in turn, stimulates new construction (SW quadrant). In the long run, this leads to a greater stock of real estate space (SE quadrant). The slopes of the different rays determine the sizes of the changes in these four variables. For instance, if construction were very inelastic with respect to real estate prices, then the new levels of rents and asset prices would be much higher than before, and new construction and the stock of space would have expanded only slightly.

Shifts in the demand for real estate assets may result from a number of factors, such as changes in short-term or long-term interest rates, the tax treatment of real estate, and the availability of construction financing, as pointed out by DiPasquale and Wheaton (1992).

If interest rates in the rest of the economy rise, then the existing yield from real estate becomes low relative to fixed-income securities, and investors will wish to shift their funds from the real estate sector; if interest rates fall, then yield from real estate becomes higher, and investors shift funds into the real estate sector. Similarly, if the risk characteristics of real estate are perceived to have changed, then the existing yield from real estate may also become insufficient (or more than necessary) to get investors to purchase real estate assets relative to other assets. Finally, changes in how real estate income is treated in the U.S. tax code can also greatly impact the demand to invest in real estate. Favorable depreciation rules for real estate (e.g., short tax life and accelerated depreciation schedule) increase the after-tax yield generated by real estate, which increases the demand to hold real estate assets. Reductions in long-term interest rates, decreases in the perceived risk of real estate, and generous depreciation or other favorable changes in the tax treatment of real estate will cause a reduction in the income that investors require from real estate.

These changes can also be examined using the four-quadrant model. As noted by DiPasquale and Wheaton (1992), higher interest rates, greater perceived risk, and adverse tax changes rotate the ray in a clockwise manner. Given a level of rent from the property market, a reduction in the current yield or capitalization rate for real estate raises asset prices, and in the SW quadrant expands construction. Eventually this increases the stock of space (in the SE quadrant), which then lowers rents in the property market for space (NE quadrant). A new equilibrium requires that the initial and finishing rent levels be equal to each other. This new equilibrium results in a new solution that is lower and more rectangular than the original. In the new equilibrium, asset prices must be higher and rents lower, while the long-term stock and its supporting level of construction must be greater. If rents were not lower, the stock would have to be the same (or lower) and this would be inconsistent with higher asset prices and greater construction. If asset prices were not higher, rents would be lower, and this would be inconsistent with the reduced stock (and less construction) that lower asset prices would generate.

As we have seen, the four-quadrant model is a useful tool for analyzing the short-run and long-run equilibrium, as well as its potential perturbations, in the property market (rent determination), the asset market (valuation and construction), and the property market (stock adjustment).

14.6 CONCLUSION

This chapter has reviewed the attributes, asset allocation, categories, and return drivers of real estate. It also presented the four-quadrant model, which allows the simultaneous assessment of the long-run equilibrium within and between the real estate space and asset markets, and provides a graphic representation of the dynamics of a real estate system.

The next four chapters will help us gain a deeper understanding of real estate as an investment in a number of fronts. Chapter 15 examines real estate indices and performance evaluation. The chapter discusses the two main approaches to indexation (appraisal-based and transaction-based), and focuses on the consequences and remedies of data smoothing in real estate. Chapter 16 examines investment styles (core, value added, and opportunistic, the main approach to categorizing investments within the category of private commercial real estate equity), portfolio allocation, and the challenges of using real estate derivatives to hedge market risk in real estate markets. Chapter 17 presents the main characteristics of unlisted (open-end funds and closed-end funds) and listed real estate products (REITs and exchange-traded funds [ETFs] based on real estate indices), as well as the extent to which analysis of publicly traded real estate securities may be used to provide information on the risks and returns of private real estate. Finally, Chapter 17 examines international real estate investments, including their potential opportunities (mainly enhanced returns, diversification benefits, potential tax advantages, and leverage), as well as the main challenges when going international (lack of local knowledge, agency costs, regulatory restrictions on foreign ownership, higher transaction costs, complex taxation, and exchange rate risk, among others).

NOTE

1. The ray would be almost vertical if construction could be supplied at any level with almost the same costs. However, land scarcity, bottlenecks, and so on, lead to an inelastic (i.e., insensitive) supply and, hence, to a ray that is more horizontal.

REFERENCES

Cohen & Steers. 2015. “About REITs.” https://www.cohenandsteers.com/insights/education/about-reits.

Deloitte. 2015. "2015 Commercial Real Estate Outlook." www2.deloitte.com/us/en/industries/real-estate.html?icid=top_real-estate.

DiPasquale, D., and W. C. Wheaton. 1992. "The Markets for Real Estate Assets and Space: A Conceptual Framework." *Journal of the American Real Estate and Urban Economics Association* 20 (1): 181–97.

Federal Reserve. 2015. "Z.1 Financial Accounts of the United States." Federal Reserve Statistical Release, December 10. www.federalreserve.gov/releases/z1/.

Geltner, D., N. Miller, J. Clayton, and P. Eichholtz. 2014. *Commercial Real Estate Analysis and Investments*. Mason, OH: OnCourse Learning.

Idzorek, T., M. Barad, and S. Meier. 2007. "Global Commercial Real Estate." *Journal of Portfolio Management* 33 (5): 37–52.

Market Reports Online. 2015. "Residential Mortgages: Global Industry Guide." January. www.marketreportsonline.com/395332.html.

National Association of Real Estate Investment Trusts. 2016. "REIT Industry Fact Sheet." January. https://www.reit.com/sites/default/files/media/PDFs/MediaBriefs/MediaFactSheetDec2015.pdf.

Savills. 2014. "UK Housing Stock Value Climbs to £5,205,000,000,000 but the Gap between the Haves and the Have Nots Grows." Savills, London, January 27. www.savills.co.uk/_news/article/72418/172125-0/01/2014/uk-housing-stock-value-climbs-to-%C2%A35-205-000-000-000-but-the-gap-between-the-haves-and-the-have-nots-grows.

CHAPTER 15

Real Estate Indices and Unsmoothing Techniques

Real estate indices are an increasingly important aspect of real estate investment management. The uses of these indices include the estimation of risks and returns for assisting the asset allocation decision-making process, as well as the specification of benchmarks for performance attribution. Performance attribution provides valuable information both for bottom-up investment management (e.g., in the selection of properties or managers) and for top-down investment management (in the determination of allocation to various categories of real estate investments).

In valuation, real estate investors often focus on the most recent sales of similar and nearby properties to form their central gauge of changes in property values. But best practices for institutional investing call for a more structured and less localized view of valuation and performance attribution. This chapter focuses on the challenges and opportunities of real estate indexation.

The two main approaches to indexation are appraisal-based and transaction-based, each of which has its own potential problems. This chapter compares these approaches and reviews many of the most popular real estate indices, which vary in terms of methodology used. The prevalence of a variety of indexation methodologies highlights the fact that all methodologies have nontrivial problems and that real estate analysts should be aware of the challenges associated with each methodology.

Using past data to infer the future risks and returns of real estate is an important part of real estate analysis, portfolio allocation, and risk management. This chapter focuses on one of the most crucial tasks of empirical analysis of real estate returns: data **unsmoothing**, which is the process of removing the effects of smoothing from a data series by reducing the level of autocorrelation in a time series. When appraisal or other backward-looking methods are used to construct real estate price indices, the resulting indices are considered to be smoothed, as the index values will slowly adjust to changes in market conditions affecting real estate prices. Unsmoothing of an index is the process of adjusting the index values so that they reflect changes in economic conditions on a timely basis. The effects of smoothed data are not limited to values based on real estate appraisals. Smoothing can affect real estate valuation besides appraisals, as well as other alternative investments, such as hedge funds and private equity funds. Therefore, the unsmoothing procedures discussed in this chapter are central to the analysis of alternative investments.

There are substantial challenges faced by real estate **appraisers** and other financial professionals who are asked to place values on an asset through time. Consider a major market movement in the midst of a more general period of financial stability. For example, suppose that the equity market experiences a general and rapid price rise of 10%. It is possible that the reported prices of asset classes that tend to be correlated with equity markets may indicate a delayed reaction to that rise and the accompanying changes in economic conditions. To the extent that a reported price or return series demonstrates a delayed response, due perhaps to the valuation methods used by appraisers or other financial professionals responsible for publishing prices, the resulting price or return series is referred to as a **smoothed series**.

Tradable prices that are smoothed can be arbitraged if transaction costs are relatively small. Nonmarket prices, such as appraisals, cannot be arbitraged, and therefore the smoothing may be more pronounced and permanent. Smoothed price and return data interfere with risk management and other financial analysis, and therefore unsmoothing of the data is essential. The first part of this chapter discusses unsmoothing of a price index or return series—the process of removing the effects of smoothing from a data series. It begins by introducing smoothed pricing and the principles of unsmoothing. The chapter also explains **transaction noise**, which arises when real estate transaction prices contain errors that make those prices less reliable when compared to prices of more liquid assets. For example, the reported transaction prices result from a negotiation process between buyers and sellers and therefore represent one set of possible values from a range of prices that would have been acceptable to both buyers and sellers. Transaction noise is another important technical issue when dealing with real estate indices. Property values are noisy (in the sense that they reflect random error) because empirical real estate values are imprecise indicators of true value. Finally, the chapter discusses the performance of various appraisal-based and transaction-based real estate indices.

15.1 SMOOTHED PRICING

Exhibit 15.1 provides a numerical illustration of the concept of smoothed pricing. Overall equity market returns, shown in column 2, are assumed to experience a 10% surge in prices in the midst of a larger time period of stable values. Three other return series are illustrated in columns 3, 4, and 5, each of which eventually responds with an 8% rise (ignoring compounding for simplicity), consistent with having true

EXHIBIT 15.1 Illustration of Price Smoothing

(1) Time Period	(2) Market Returns	(3) Unsmoothed	(4) Lightly Smoothed	(5) Strongly Smoothed
1	0%	0%	0%	0%
2	10%	8%	6%	4%
3	0%	0%	2%	4%
4	0%	0%	0%	0%
Mean	2.5%	2%	2%	2%
Std. dev.	5.0%	4.0%	2.8%	2.3%
Corr. w/mkt.	1.00	1.00	0.94	0.58
Beta	1.00	0.80	0.53	0.27

underlying (unsmoothed) betas of 0.80. The unsmoothed return series experiences its entire price response (8%) to the market in the same time period as the equity market. The lightly smoothed return series reflects 75% of its 8% price response in the same period as the market, but reflects another 25% of its price response in the subsequent time period. The strongly smoothed return series experiences half of its price response in the same period as the market, and the other half in the subsequent period.

15.1.1 Price Smoothing and Arbitrage in a Perfect Market

In a perfect market (i.e., one without transaction costs or trading restrictions), arbitrageurs should exploit profit opportunities caused by the consistently delayed price responses contained in smoothed prices. Any asset with consistently delayed price responses will be purchased after general prices rise and will be short-sold by arbitrageurs after general prices decline. In this way, arbitrageurs will consistently profit from delayed price increases using long positions, and profit from delayed price declines using short positions. For example, an arbitrageur could buy the strongly smoothed asset at the end of time period 2 (immediately after the market rises) and expect, on average, to receive an alpha of 4% in time period 3, as the asset's price experiences a delayed response to the large market rise in time period 2. In the case of a large market decline in a perfect market, the arbitrageur can short-sell a smoothed price immediately after a large decline in the prices of similar assets that are not smoothed. Note that the arbitrageur can hedge risk by taking offsetting positions in similar assets that are not smoothed.

Competition both to buy assets with smoothed prices before delayed price increases and to short-sell assets with smoothed prices before delayed price declines will drive away delayed prices in tradable assets by forcing prices to respond more quickly. Eventually, in a perfect market, competition among arbitrageurs will force prices to respond fully and immediately in the absence of transaction costs. For example, in the case of a large rise in unsmoothed prices, arbitrageurs will vie to establish positions earlier and earlier in the anticipation of subsequent price movements. Arbitrageurs will force previously smoothed prices to respond to price changes as quickly as they are reflected in unsmoothed prices. Thus, in perfect markets, any smoothing of price returns is unsmoothed by the actions of arbitrageurs whenever assets can be traded at the stated prices.

15.1.2 Persistence in Price Smoothing

There are two primary culprits that prevent smoothed return series from being unsmoothed by arbitrageurs. First, the return series may not indicate true trading opportunities. Appraisals, for example, are typically indications of prices that do not represent either bids to buy or offers to sell. Appraisals are used to value portfolios for accounting purposes and to construct price indices. In both cases, the appraisals represent estimated values, not market prices.

Second, even if a smoothed return series indicates trading opportunities (i.e., prices at which transactions may be made), the underlying assets may have substantial transaction costs or other barriers to arbitrage. For example, in real estate, the time and transaction costs of buying and selling assets in order to exploit delayed pricing responses may be prohibitively expensive relative to the potential gains from moderate price smoothing. Real estate sales commissions, real estate transfer taxes, legal costs, financing costs, search costs, inspection costs, and so forth, provide substantial barriers to arbitrageurs seeking to exploit the lags in price changes caused by price smoothing. Limited partnership holdings in real estate and private equity funds may have a stated life, such as 10 years. During this period, liquidation of fund interests is generally infeasible until the general partner exits the underlying investments.

In addition to transaction costs, there are other barriers to trading assets that exhibit smoothed pricing. For example, international open-end equity mutual funds were notorious for allowing stale prices to cause smoothing in their reported prices (net asset values). Arbitrageurs exploited the smoothed pricing by establishing long positions in international funds when domestic markets rose sharply, and establishing neutral positions in international funds when domestic markets declined (and when foreign markets were already closed). Many mutual fund companies implemented more accurate pricing methods or erected powerful barriers against short-term trading of such funds, such as a 2% redemption fee on positions held for less than 90 days.

Due to the actions of arbitrageurs and other market participants, assets with tradable prices, low transaction costs, and minimal trading barriers do not typically require unsmoothing. The need to unsmooth prices tends to be greater for nontradable prices and assets with high transaction costs or trading barriers.

15.1.3 Problems Resulting from Price Smoothing

The smoothing of the last two return series in Exhibit 15.1 generates lower standard deviations, lower correlations with the market, and much lower reported betas. For example, note that the true beta of 0.80 is indicated in the estimated beta of the unsmoothed return series, but the lightly smoothed return series has a reported beta one-third smaller than the unsmoothed return series, and the strongly smoothed return series has a reported beta two-thirds smaller. Similarly, the standard deviations of the smoothed series are substantially lower, as smoothing causes the largest outliers of the unsmoothed series to be muted. The primary problem resulting from price smoothing is that it causes substantial understatement of both volatility and correlation.

Risk understatement may cause inappropriately high allocations to assets with smoothed prices. Portfolio optimization models will tend to overweight assets with understated risk. Furthermore, underestimated price correlations due to price smoothing may distort the estimation of appropriate hedge ratios and interfere with risk management.

Though risk can be understated through price smoothing, long-term historical mean returns are not substantially affected by the price-smoothing process. Nevertheless, investors selecting investments with high Sharpe ratios will be attracted to asset classes with smoothed prices, as the mean return in the numerator of the Sharpe ratio is typically only slightly affected by smoothing, whereas the denominator contains artificially low estimates of standard deviation. Smoothing results in artificial increases in the estimated Sharpe ratio and other similar performance measures.

15.2 MODELS OF PRICE AND RETURN SMOOTHING

In order to detect, correct, or exploit smoothing, it is necessary to understand how smoothing takes place and how it can be modeled. A properly specified model can be used to determine a method for estimating unsmoothed prices or returns. This section discusses the primary approaches to modeling smoothing.

15.2.1 Reported Prices as Lags of True Prices

Define $P_{t,\text{reported}}$ as the reported or smoothed price of an asset at time t, and $P_{t,\text{true}}$ as the true price. An example of a reported price would be a price index based on appraisal values or a hedge fund's net asset value that is subject to smoothing by a fund manager. The true price of the asset is defined as the best indication of the market price at which the asset would trade with ready buyers and sellers.

For example, consider a real estate index in which at least one of two things happens: (1) the prices contained in the index partially represent lagged market values, or (2) the prices in the index contain the professional judgment of appraisers who based their opinions at least partially on a series of lagged market values. In either case, current reported prices are a function of past true prices. For example, Equation 15.1 contains a very general model of smoothing that expresses the reported price as a moving average of the current true price and past true prices:

$$P_{t,\text{reported}} = \alpha + \beta_0 P_{t,\text{true}} + \beta_1 P_{t-1,\text{true}} + \beta_2 P_{t-2,\text{true}} + \cdots \qquad (15.1)$$

Equation 15.1 allows the reported and smoothed price index at time t ($P_{t,\text{reported}}$) to depend not just on the contemporaneous true price ($P_{t,\text{true}}$) but also on the true prices in previous time periods. The relationships between the current reported price and the true prices are specified using a set of parameters denoted as β_i. A potentially simpler model that has only one parameter specifies an exact relationship between the parameters in Equation 15.1, as shown in Equation 15.2:

$$P_{t,\text{reported}} = \alpha P_{t,\text{true}} + \alpha(1-\alpha)P_{t-1,\text{true}} + \alpha(1-\alpha)^2 P_{t-2,\text{true}} + \cdots \qquad (15.2)$$

where α is a parameter greater than zero and less than or equal to 1 that determines the speed of the decay function. A decay function is simply a numeric construct that puts less weight on older valuations and more weight on more recent valuations. Consider the case of $\alpha = 0.50$ in Equation 15.2. In this case, the current reported price depends 50% on the current true price, 25% on the true price of the previous observation date, 12.5% on the true price of the observation date from two periods before, and so on. Put differently, the true price in a particular period is only 50% reflected in the reported price of the same time period, 25% reflected on a one-period-delayed basis, 12.5% reflected on a two-period-delayed basis, and so on. A value of $\alpha = 0.60$ places 60% weight on the current true price, 24% weight on the immediately prior true price, and so on. The weights of the prices in Equation 15.2 sum to 1. A value of $\alpha = 1.00$ indicates that true prices are immediately and fully reflected in reported prices. A value of α approaching zero indicates that the effect of a true price on reported prices occurs on a more delayed basis.

Equation 15.2 can be factored to generate a highly simplified expression for true price as a function of the current reported price and the one-period-lagged value of the reported price, as shown in Equation 15.3:

$$P_{t,\text{true}} = (1/\alpha) \times P_{t,\text{reported}} - [(1-\alpha)/\alpha] \times P_{t-1,\text{reported}} \tag{15.3}$$

The importance of Equation 15.3 is that it expresses the most recent true but unobservable price as a simple equation involving the most recent smoothed index value and the previous smoothed index value, both of which are observable. Equation 15.4 rearranges Equation 15.3 into a potentially more intuitive form:

$$P_{t,\text{true}} = P_{t-1,\text{reported}} + [1/\alpha \times (P_{t,\text{reported}} - P_{t-1,\text{reported}})] \tag{15.4}$$

Equation 15.4 indicates that the true price differs from the previously reported price by an amount based on the most recent price change in the reported price series. Consider the case of a rising reported value $P_{t,\text{reported}} > P_{t-1,\text{reported}}$, and note that in Equation 15.4, the expression $1/\alpha$ is greater than 1 (assuming that α is between zero and 1). In this case, the true value of the asset is expressed as the previously reported value of the asset ($P_{t-1,\text{reported}}$) plus the reported price change increased by a factor of $1/\alpha$. For example, with $\alpha = 0.60$, a \$10 change in the reported price implies a \$16.67 difference between the current true price and the previous reported price. The muted price change of \$10 in the reported (and smoothed) index is consistent with a much larger change in the true underlying price.

The primary importance of Equation 15.4 is that given an estimate of the parameter α, the equation can be used to generate the estimated true prices and their changes (i.e., the unsmoothed prices) from a series of smoothed prices. Fisher (2005) estimates a value of $\alpha = 0.40$ for private unleveraged annual real estate returns in the United States. Inserting $\alpha = 0.40$ into Equation 15.4 indicates that true prices should be estimated based on a price change that is 2.5 times larger than the most recent reported price change.

15.2.2 Modeling Lagged Returns Rather Than Prices

Section 15.2.1 focused on prices. Often the relationship is specified directly in terms of returns. Generally, there is a small difference between imposing the lagged structure specified by Equation 15.2 on prices and imposing the same structure on returns.[1] But as an approximation, Equation 15.1 can be written in terms of returns, as in Equation 15.5:

$$R_{t,\text{reported}} \approx \beta_0 R_{t,\text{true}} + \beta_1 R_{t-1,\text{true}} + \beta_2 R_{t-2,\text{true}} + \cdots \tag{15.5}$$

where $R_{t,\text{reported}}$ is the return on the reported price series in period t, and $R_{t,\text{true}}$ is the return on the true but unobservable price series in period t.

15.2.3 Estimating the Parameter for First-Order Autocorrelation

Equation 15.3 indicates that true prices can be determined from reported prices using the parameter α. Accordingly, estimated values of true prices can be determined from reported prices using an estimation of the parameter.

An intuitive interpretation of α is that it determines the extent to which the reported price (or return) in a particular time period is determined or driven by the value of the true price (or return) in the same time period. The higher the value of α, the more the current reported value is driven by current changes in the true value rather than by past changes.

Equation 15.6 is formed by multiplying each side of Equation 15.4 by α, substituting $\alpha = 1 - \rho$, and rearranging the terms:

$$P_{t,\text{reported}} = (1 - \rho)P_{t,\text{true}} + \rho P_{t-1,\text{reported}} \tag{15.6}$$

Equation 15.6 can be expressed as an approximation in terms of price changes and returns, as depicted in Equations 15.7 and 15.8:

$$\Delta P_{t,\text{reported}} \approx (1 - \rho)\Delta P_{t,\text{true}} + \rho\Delta P_{t-1,\text{reported}} \tag{15.7}$$

$$R_{t,\text{reported}} \approx (1 - \rho)R_{t,\text{true}} + \rho R_{t-1,\text{reported}} \tag{15.8}$$

where ρ is the first-order autocorrelation coefficient, given the assumption that the reported price series (or return series) is autoregressive of order one (Gallais-Hamonno and Nguyen-Thi-Thanh 2007).[2] The use of ρ in place of α puts the emphasis on a statistical interpretation of the relationships.

Equation 15.8 expresses first-order autocorrelation in the reported return series because the smoothed return in period t (the left-hand side of the equation) is correlated by ρ with the smoothed return in the immediately preceding time period (the last term on the right-hand side). The most current return of the smoothed return series depends partially on the new market information contained in the true return ($R_{t,\text{true}}$) and partially on the smoothed return of the previous time period. The parameter ρ specifies the relative importance of the two explanatory variables, with higher values of ρ indicating greater smoothing.

For example, assume that ρ in Equation 15.8 is equal to 40% and that the true underlying price changes in a particular asset class in the last five time periods were 0%, 0%, 10%, 0%, and 0%. Applying Equation 15.8 to obtain a series of smoothed prices would generate the following return series: 0%, 0%, 6%, 2.4%, and 0.96%. Note that the smoothed return for time period 4 is 2.4% rather than 4.0%. The reason for this is that in Equation 15.8, ρ is multiplied by the lagged smoothed return (6%), not the lagged true return (10%). Similarly, the impact of the 10% true return on the smoothed returns will continue to be experienced forever, but with a rapidly diminishing effect.

15.2.4 Four Reasons for Smoothed Prices and Delayed Price Changes in an Index

There are four primary explanations for first-order autocorrelation in a price index, such as an index of real estate prices. One explanation is that a price index is being based on observed prices of the most recent transactions of each component of the index and that old or stale prices are being used for index components that have not recently traded. Let's return to the previous example, in which $\rho = 40\%$ and a series of true underlying price changes in a particular asset class in the last five time periods is 0%, 0%, 10%, 0%, 0%. In the case of a **transaction-based price index**, Equation 15.8 and $\rho = 40\%$ would be consistent with the idea that 60% of the most recent underlying asset values were based on a transaction that occurred in the current time period (and reflected the true return of 10% in period 3), while the impact of the 10% true return on the remaining 40% of the underlying asset values will be reflected through time as they transact in subsequent time periods (as reflected in period 4 and period 5 returns of 2.4% and 0.96%).

A professional appraiser may also generate a series of smoothed prices. One reason is that the appraiser observes price changes on a delayed basis and only on those properties that transacted. Another reason is that the appraiser may exhibit the behavioral phenomenon known as anchoring. Anchoring is the observed tendency of humans to give disproportionate weight or reliability to previous observations. In the previous example, an appraiser may be reluctant to believe that underlying assets have truly risen by 10%, and may move valuations 60% in the direction of 10% during the first period, and then continue to adjust valuations in the subsequent periods.

A third major reason for smoothed pricing is that even current transaction prices in an efficient market may be selected such that they signal lagged price responses. Consider an efficient real estate market with two types of properties of approximately equal total market value. Assume that one of the property types experiences a 5% true price growth, while the other type experiences a 25% true price growth. Based on the assumption of equal weights, the true total real estate price index should indicate a 15% price increase. However, in a rapidly rising market of real estate prices, investors may be systematically biased toward transacting in those property types with characteristics that caused lower price growth (5%). The primary explanation for the tendency of market transactions to be dominated by assets that rise less in price during a bull market may be behavioral, such as the reluctance of buyers to accept the 25% price increase of the other property type. Simply put, more buyers are willing to buy property types with 5% higher prices than are willing to buy property types with 25% higher prices, even when the higher prices of each property type reflect true market values. It should be noted that this is only one of many scenarios that could occur.

Continuing the example, if 80% of the properties that transact are of the type that experienced lower price growth (5%), then the average price changes of the observed transactions would be 9%, found as $[(80\% \times 5\%) + (20\% \times 25\%)]$. The 9% average price index change is 60% of the true price change (15%) for the entire market. Presumably, buyers will increasingly buy the property types that have risen in price by 25% once the new and higher price level becomes familiar. Similarly, price declines will be reported on a smoothed basis if sellers are reluctant to sell the property types that have declined the most.

The final reason for smoothed prices is the potential delay between the setting of a price on a real estate transaction and the reporting of the transaction. A real estate price may be negotiated months before the transaction occurs, and the reported price of the transaction may become known to the appraiser or index on a delayed basis as well.

15.3 UNSMOOTHING A PRICE OR RETURN SERIES

The previous section discussed smoothing and provided simple examples of how smoothed prices are formed through delayed responses to true price changes. But the objective in practice is to estimate true returns from smoothed returns. This section discusses unsmoothing: the process of estimating a true but unobservable price or return series from an observable but smoothed price or return series.

15.3.1 Unsmoothing First-Order Autocorrelation Given ρ

Equation 15.8 can be factored by solving for the true return based on the reported or smoothed returns, as shown in Equation 15.9 (and, for expositional simplicity, replacing the approximation sign with an equality sign):

$$R_{t,\text{true}} = (R_{t,\text{reported}} - \rho R_{t-1,\text{reported}})/(1-\rho) \tag{15.9}$$

Returning to the example of the true return series used in the previous section (0%, 0%, 10%, 0%, and 0%) to generate the smoothed return series of 0%, 0%, 6%, 2.4%, and 0.96%, Equation 15.9 can be used along with $\rho = 0.40$ to back out the true return series from the reported series. Inserting the smoothed return series as the reported returns in Equation 15.9 and continuing to use $\rho = 0.40$, the implied true returns are as follows:

$$R_{3,\text{true}} = (R_{3,\text{reported}} - \rho R_{2,\text{reported}})/(1-\rho) = [6\% - (0.40 \times 0\%)]/(1-0.40) = 10\%$$
$$R_{4,\text{true}} = (R_{4,\text{reported}} - \rho R_{3,\text{reported}})/(1-\rho) = [2.4\% - (0.40 \times 6\%)]/(1-0.40) = 0\%$$
$$R_{5,\text{true}} = (R_{5,\text{reported}} - \rho R_{4,\text{reported}})/(1-\rho) = [0.96\% - (0.40 \times 2.4\%)]/(1-0.40) = 0\%$$

Note that the smoothed return series can be used without error to find the underlying true return series if the process follows a first-order autocorrelation process without an error term and if ρ can be estimated without error. Of course, in practice, the returns do not conform perfectly to the first-order autocorrelation model, and the coefficient ρ must be estimated and is subject to estimation error. Thus, the unsmoothed estimations contain errors.

APPLICATION 15.3.1

A smoothed return series has returns in two consecutive time periods of −2% and +10%. The estimated autocorrelation is 0.60. Find the estimated true return of the series in the second period. Inserting the two returns and the estimated autocorrelation coefficient into Equation 15.9 produces the following equation:

$$R_{2,\text{true}} = [10\% - (0.60 \times -2\%)]/(1-0.60)$$
$$R_{2,\text{true}} = (10\% + 1.2\%)/0.4 = 28\%$$

It would take a very high true return in the second period (i.e., 28%) to generate a smoothed return of 10% after a smoothed return of −2%. Note that the equation contains four variables. Given any three of the variables, the equation can be used to solve for the fourth variable.

More complex models may be appropriate when the smoothing takes place over more than one time period, such as a fourth-order autocorrelation model, when smoothing of quarterly returns takes place over a one-year period. The following section discusses the three-step process for unsmoothing a reported price or return series to estimate a true price or return series.

15.3.2 The Three Steps of Unsmoothing

Unsmoothing a return series containing autocorrelation involves three steps:

STEP 1: The first step is determining or specifying the form of the autocorrelation. As throughout this chapter, first-order autocorrelation is being assumed (as shown in Equation 15.8).

STEP 2: The second step is estimating the parameter(s) of the assumed autocorrelation process. In our example of a first-order autocorrelation coefficient, the only parameter is ρ, as indicated in Equations 15.8 and 15.9. The first-order autocorrelation coefficient of a series is found as the correlation coefficient between each observation and the observation from the same series in the previous time period, as depicted in Equation 15.10:

$$\hat{\rho} = \text{corr}(R_{t,\text{reported}}, R_{t-1,\text{reported}}) \tag{15.10}$$

Equation 15.11 depicts the formula for a correlation coefficient based on covariance and standard deviations:

$$\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j) \tag{15.11}$$

where ρ_{ij} is the correlation coefficient between two variables, σ_{ij} is the covariance between the two variables, and σ_i and σ_j are the standard deviations of the two variables. The correlation coefficient between each observation and its value in the previous time period can be estimated using sample statistics. Let's return to the example of the smoothed return series of 0%, 0%, 6%, 2.4%, and 0.96%. Using Equation 15.11, the estimated correlation coefficient between that series and the one-period lagged return series of 0%, 0%, 0%, 6%, and 2.4% is 0.037. Note that the estimated value of ρ (0.037) is far from its assumed and true value (0.40). The explanation for the difference is the small sample size and estimation error.

STEP 3: The third step is inserting the estimated correlation coefficient in place of ρ in Equation 15.9 and solving for $R_{t,\text{reported}}$. Using Equation 15.9 and the estimated correlation coefficient (0.037) rather than the true correlation coefficient (0.40) generates the series 0%, 0%, 6.2%, 2.3%, and 0.9%:

$$R_{2,\text{true}} = [0\% - (0.037 \times 0\%)]/(1 - 0.037) = 0\%$$
$$R_{3,\text{true}} = [6\% - (0.037 \times 0\%)]/(1 - 0.037) = 6.2\%$$
$$R_{4,\text{true}} = [2.4\% - (0.037 \times 6\%)]/(1 - 0.037) = 2.3\%$$
$$R_{5,\text{true}} = [0.96\% - (0.037 \times 2.4\%)]/(1 - 0.037) = 0.9\%$$

The smoothed return series for time periods 2 through 5 (0%, 6%, 2.4%, and 0.96%) is unsmoothed as 0%, 6.2%, 2.3%, and 0.9%, slightly closer to the assumed true return series (0%, 10%, 0%, and 0%) from which the example was derived. The reason for the very limited success was the poor estimation of ρ (0.037 as an estimation of 0.40). As was demonstrated in section 15.3.1, the use of $\rho = 0.40$ to unsmooth the return series generates the exact true series. The success of the unsmoothing therefore depends on the proper specification of the autocorrelation scheme and especially the accurate estimation of the parameter(s). It would be expected that the estimation of ρ should improve as the sample size is increased and that the poor estimation of ρ in the example was attributable to the use of such a small sample.

15.3.3 Unsmoothing Using Prices Rather Than Returns

As indicated in Equation 15.8, the unsmoothing of returns is an approximation if the true relationship is based on prices. The process illustrated for returns throughout this chapter can easily be performed using price changes and Equation 15.7, even if the data are provided as a series of returns. The six steps to unsmoothing a return index based on a model of smoothed price changes (Equation 15.7) are:

1. Convert the returns to a price index using a cumulative wealth index that includes compounding.
2. Convert the price index to a series of price changes.
3. Apply Equation 15.11 to estimate the correlation between the price change series and its lagged value.
4. Apply Equation 15.9, substituting price changes for returns.
5. Use the unsmoothed price changes to form a price index.
6. Convert the unsmoothed price index back into returns.

15.3.4 Unsmoothing Returns with Higher-Order Autocorrelation

The previous sections discussed first-order autocorrelation, wherein the return of the smoothed series is fully explained by the true return in the same period and the smoothed return from the previous time period. More advanced analyses of return autocorrelation allow for the current true return or price to depend on the previous reported or smoothed returns or prices of two or more previous periods on a more general basis than first-order autocorrelation. For example, let's assume that lagged effects up to k periods generate Equation 15.12:

$$R_{t,\text{reported}} = \alpha + \beta_1 R_{t-1,\text{reported}} + \beta_2 R_{t-2,\text{reported}} + \cdots + \beta_k R_{t-k,\text{reported}} \qquad (15.12)$$

Note that the effect of the true return in time period t on $R_{t,\text{reported}}$ is assumed to be captured by the intercept and error term, since true returns in period t are assumed to be uncorrelated with previous returns.

The important distinction between Equation 15.12 and first-order autocorrelation (Equation 15.9) is that first-order autocorrelation specifies the exact relationship between coefficients (i.e., βs) and reduces the coefficients to being specified by one parameter (ρ). Equation 15.12 provides a more general (flexible) specification by allowing the current reported return to depend on returns of various lags without constraining their relationships to each other.

Finally, note the case of $k = 1$ in Equation 15.12, which generates the simple linear regression coefficient between $R_{t,\text{reported}}$ and $R_{t-1,\text{reported}}$. A useful formula for a simple regression coefficient from regressing variable y_t on variable x_t is that $\beta = \rho_{xy}\sigma_y/\sigma_x$. Note that in the case of regressing a variable on its lagged value, the true standard deviation of each series is equal, so that $\sigma_x = \sigma_y$ and $\beta = \rho_{xy}$ (and the estimated standard deviations of the original and lagged series approach each other as the sample size increases). Thus, the true slope coefficient from a linear regression between a variable and its lagged value is equal to the correlation coefficient, which in turn is equal to the first-order autocorrelation coefficient. Therefore, the first-order autocorrelation assumed throughout the examples in this chapter is the case of $k = 1$ in the more general model shown in Equation 15.12.

15.4 AN ILLUSTRATION OF UNSMOOTHING

The purpose of this section is to provide a detailed example of the unsmoothing of an actual return series with first-order autocorrelation. This section follows the three-step procedure discussed in section 15.3.2.

15.4.1 The Smoothed Data and the Market Data

Exhibit 15.2 contains eight years of quarterly return data from two popular U.S. real estate indices. The time period of analysis is shortened to 32 quarters in this analysis so that all of the data can be shown in a concise exhibit. Column 3 contains returns based on the National Council of Real Estate Investment Fiduciaries (NCREIF) Property Index (henceforth NPI), discussed in detail later in the chapter. The NPI is based on appraised prices of private real estate properties and is therefore likely to contain substantial price smoothing. Column 4 contains returns based on the all-equity REIT index of the FTSE National Association of Real Estate Investment Trusts (NAREIT) U.S. Real Estate Index Series (henceforth NAREIT Index), which is based on closing market prices of publicly traded equity real estate investment trusts (REITs). To the extent that the market for REITs is informationally efficient, the return series should not have autocorrelation. The NAREIT Index serves in the example as a proxy of a true return series (even though it contains transaction noise, as will be discussed in section 15.5), and the NPI serves as a proxy of a smoothed return series. Though both series have relatively similar mean returns, the standard deviation of returns based on the NPI is substantially lower than that of the NAREIT Index, as shown in Exhibit 15.2. Part of the volatility difference can be explained by the fact that the NPI reflects no leverage (i.e., reflects underlying real estate assets without leverage), whereas the NAREIT Index series reflects the returns of REITs that generally reflect levered real estate positions.

Note the very large negative returns of the NAREIT Index in the fourth quarter of 2008 and the first quarter of 2009. Note further that the appraised series (the NPI) shows only relatively modest negative returns in the same two quarters, apparently reflecting the tendency of appraised prices to move in only partial response to true market price changes (although the muted reaction can also be explained at least in part by the lack of leverage in the NPI relative to the NAREIT Index).

Note also that the NAREIT Index has very large gains in the next two quarters (the second and third quarters of 2009). However, the NPI continues to drift

EXHIBIT 15.2 Unsmoothing of Quarterly Returns, 2007–14

		Returns		Lagged Values		Unsmoothed Values	
Year	Quarter	NPI	NAREIT	NPI	NAREIT	NPI	NAREIT
2007	1	3.62%	3.46%				
2007	2	4.59%	−9.40%	3.62%	3.46%	9.87%	−12.89%
2007	3	3.56%	2.59%	4.59%	−9.40%	−2.04%	5.84%
2007	4	3.21%	−1.27%	3.56%	2.59%	1.31%	−2.31%
2008	1	1.60%	1.40%	3.21%	−1.27%	−7.16%	2.12%
2008	2	0.56%	−4.93%	1.60%	1.40%	−5.10%	−6.65%
2008	3	−0.17%	5.55%	0.56%	−4.93%	−4.14%	8.39%
2008	4	−8.29%	−38.80%	−0.17%	5.55%	−52.47%	−50.83%
2009	1	−7.33%	−31.87%	−8.29%	−38.80%	−2.11%	−29.99%
2009	2	−5.20%	28.85%	−7.33%	−31.87%	6.39%	45.33%
2009	3	−3.32%	33.28%	−5.20%	28.85%	6.91%	34.48%
2009	4	−2.11%	9.39%	−3.32%	33.28%	4.47%	2.91%
2010	1	0.76%	10.02%	−2.11%	9.39%	16.38%	10.19%
2010	2	3.31%	−4.06%	0.76%	10.02%	17.19%	−7.88%
2010	3	3.86%	12.83%	3.31%	−4.06%	6.85%	17.41%
2010	4	4.62%	7.43%	3.86%	12.83%	8.76%	5.96%
2011	1	3.36%	7.50%	4.62%	7.43%	−3.50%	7.52%
2011	2	3.94%	2.90%	3.36%	7.50%	7.10%	1.65%
2011	3	3.30%	−15.07%	3.94%	2.90%	−0.18%	−19.95%
2011	4	2.96%	15.26%	3.30%	−15.07%	1.11%	23.49%
2012	1	2.59%	10.49%	2.96%	15.26%	0.58%	9.20%
2012	2	2.68%	4.00%	2.59%	10.49%	3.17%	2.24%
2012	3	2.34%	1.03%	2.68%	4.00%	0.49%	0.22%
2012	4	2.54%	3.11%	2.34%	1.03%	3.63%	3.67%
2013	1	2.57%	8.10%	2.54%	3.11%	2.73%	9.45%
2013	2	2.87%	−2.13%	2.57%	8.10%	4.50%	−4.91%
2013	3	2.59%	−2.61%	2.87%	−2.13%	1.07%	−2.74%
2013	4	2.53%	−0.17%	2.59%	−2.61%	2.20%	0.49%
2014	1	2.74%	8.52%	2.53%	−0.17%	3.88%	10.88%
2014	2	2.91%	7.13%	2.74%	8.52%	3.84%	6.75%
2014	3	2.63%	−2.48%	2.91%	7.13%	1.11%	−5.09%
2014	4	3.04%	11.54%	2.63%	−2.48%	5.27%	15.34%
Arithmetic Mean		1.53%	2.55%	1.48%	2.26%	1.36%	2.59%
Std. Dev.		3.27%	13.73%	3.31%	13.86%	11.36%	17.34%
Autocorrelation of NPI			84.48%				
Autocorrelation of NAREIT			21.34%				

Source: www.reit.com and www.ncreif.org/property-index-returns.aspx.

downward throughout 2009, possibly continuing to reflect the previous declines on a lagged basis. Investors generally find it impossible to sell short private real estate at appraisal-based prices in order to take advantage of these smoothed returns, since private funds are typically not available for short selling. During a time of market turmoil, it is also difficult to sell funds of private real estate investments, as many open-end funds either cease redemptions or erect gates. Statistical analysis, detailed in the following sections, can provide objective indications of these tendencies. The subsequent sections assume first-order autocorrelation between returns.

15.4.2 Estimating the First-Order Autocorrelation Coefficient

In Exhibit 15.2, the returns to the NCREIF and NAREIT indices for the first quarter of 2007 are 3.62% and 3.46%, respectively. The same returns, when lagged, are located in the row for the second quarter of 2007. The lagged time series, with 31 quarters of data, has one less observation than the unlagged time series, meaning that their estimated standard deviations will differ, even though they are based on mostly the same data.

The formula for the correlation coefficient of a sample is used to compute the first-order autocorrelation of each series. The appraisal-based NPI has an estimated autocorrelation coefficient of 84.5%, whereas the market-price-based NAREIT Index has an estimated autocorrelation coefficient of 21.3%. It should be noted that the sample period covers the highly unusual real estate market collapse that coincided with the financial crisis that began in 2007. Accordingly, the observed correlations may not be representative of more normal economic conditions due to the presence of outliers and their potentially disproportionate influence.

15.4.3 Unsmoothing the Smoothed Return Series Given Rho (ρ)

Columns 7 and 8 of Exhibit 15.2 unsmooth the original return data in the first and second data columns using the estimated autocorrelation coefficients (84.5% and 21.3%, respectively). The unsmoothed returns are computed using Equation 15.9.

For example, the first unsmoothed return for the NPI is 9.87%, using Equation 15.9 and unrounded numbers:

$$R_{t,\text{true}} = (R_{t,\text{reported}} - \rho R_{t-1,\text{reported}})/(1 - \rho)$$
$$9.87\% = [4.59\% - (0.845 \times 3.62\%)]/(1 - 0.845)$$

Note in the case of the NPI that relatively small changes in the returns between two adjacent time periods in the smoothed series often generate large changes in the unsmoothed returns. For example, the −8.29% smoothed return in the fourth quarter of 2008 generates a massive −52.47% decline in the unsmoothed return for the same quarter. However, the −7.33% smoothed return in the next quarter (the first quarter of 2009) generates only a −2.11% decline in the unsmoothed return for the same quarter. The unsmoothing technique captures the likelihood that the second large negative return (−7.33% in the first quarter of 2009) was a lagged reaction to the events of the fourth quarter of 2008 due to smoothing.

15.4.4 The Relationship between the Variances of True and Reported Returns

In addition to calculating unsmoothed returns, the estimated autocorrelation can be used to calculate the volatility and the beta of unsmoothed return series. Bearing in mind that

$$R_{t,\text{reported}} = \rho R_{t-1,\text{reported}} + (1 - \rho) R_{t,\text{true}}$$

one can express the variance of the smoothed return as

$$\sigma^2(\text{Reported}) = \rho^2 \times \sigma^2(\text{True}) + (1 - \rho)^2\sigma^2(\text{Reported}) \tag{15.13}$$

Equation 15.13 is obtained under the assumption that the variance of the true return series and the autocorrelation of the smoothed series are both constant through time. Using some algebra will allow us to express the variance of the unsmoothed (true) return series as

$$\sigma^2(\text{True}) = \sigma^2(\text{Reported}) \times \frac{1 + \rho}{1 - \rho} \tag{15.14}$$

Example: Suppose the annual variance of a smoothed return series is 0.03 per year. The estimated autocorrelation is 22%. What is the variance of the true or unsmoothed return?

$$\sigma^2(\text{True}) = 0.03 \times \frac{1 + 0.22}{1 - 0.22} = 0.047$$

As expected, the variance of the unsmoothed series is higher than the variance estimated using the smoothed series.

APPLICATION 15.4.4

The variance of a true return series is constant at 0.280. A smoothed series of those returns has an estimated autocorrelation coefficient of 0.75 that is assumed constant. What is the estimated variance of the reported (smoothed) return series? Using Equation 15.14, the solution can be found by inserting values and solving for σ^2 (reported):

$0.280 = \sigma^2(\text{Reported})\ (1 + 0.75)/(1 - 0.75)$
$\sigma^2(\text{Reported}) = 0.280(1.75/0.25) = 0.280/(1.75/0.25) = 0.04$

Note that given any two of the three variables in Equation 15.14, it should be possible to estimate the value of the third variable. Also, it would be common for return dispersions to be discussed using standard deviation rather than variance.

15.4.5 The Relationship between the Betas of True and Reported Returns

Using an approach similar to that from the previous subsection regarding variances, one is able to present the beta of the unsmoothed return series as a function of the beta of the smoothed series. For example, if the market beta of a smoothed series is estimated using the following one-factor regression, then one can obtain an estimate of the beta of the true return series using the value of autocorrelation:

$$R_{t,\text{reported}} = \alpha_{\text{reported}} + \beta_{\text{reported}} \times R_{t,m} + \varepsilon_t$$

The beta of the true return series is

$$\beta_{\text{true}} = \frac{\beta_{\text{reported}}}{1 - \rho} \tag{15.15}$$

According to Equation 15.15, if the autocorrelation is positive ($\rho > 0$), then the true beta will be greater than the estimated beta. Therefore, the systemic risk of a smoothed return series tends to be underestimated.

Example: The market risk of a smoothed return series is estimated to be 0.7. What is the estimate of the beta of the true return series if the autocorrelation is estimated to be 0.3?

Using Equation 15.15, we have

$$\beta_{\text{true}} = \frac{0.7}{1 - 0.3} = 1$$

The beta of the true return series is estimated to be 1.0.

APPLICATION 15.4.5

The smoothed returns of a return series with an autocorrelation of 0.40 have a true beta of 1.25 after adjusting for smoothing. What is the estimated reported beta (i.e., the beta of the smoothed returns)? Inserting the true beta and autocorrelation coefficient into Equation 15.15 generates the following equation:

$$1.25 = \beta_{\text{reported}}/(1 - 0.40)$$
$$\beta_{\text{reported}} = 1.25 \times 0.60 = 0.75$$

The smoothed series exhibits a much lower beta than the true return series. Note that given any two of the three variables in Equation 15.15, it should be possible to estimate the value of the third variable.

15.4.6 Interpreting the Results of Unsmoothing

Exhibit 15.2 lists the standard deviations of both the unsmoothed and the smoothed return series. Note that the standard deviation of the smoothed NPI (3.27%) increases to 11.36% when the returns are unsmoothed. To the extent that the return series has been properly and accurately unsmoothed to reflect true values, the true volatility of the asset is more than three times the volatility perceived based on smoothed values. Note also that the NAREIT Index contains positive autocorrelation and that the estimated standard deviation of the unsmoothed NAREIT Index (17.34%) is modestly higher than the estimated standard deviation of the original NAREIT Index (13.73%). It is possible that the positive autocorrelation of the market prices contained in the NAREIT Index is a spurious outcome of the incredible turmoil of the real estate market during the financial crisis that began in 2007.

The original NPI exhibited volatility almost one-quarter that of the original NAREIT Index. The unsmoothed NPI exhibited volatility approximately two-thirds that of the unsmoothed NAREIT Index and more than three-quarters that of the original NAREIT Index. These values appear somewhat in line with the higher risk of the assets underlying the NAREIT Index due to the use of leverage underlying its assets. The dramatic increase in estimated risk for the NPI that results from unsmoothing the actual data highlights the importance of unsmoothing. Asset allocations based on volatilities of the smoothed data would dramatically overweight assets with smoothed returns in a mean-variance optimization framework. Improved unsmoothing of the return data in Exhibit 15.2 might be attained using autocorrelation techniques more general than the first-order autocorrelation model illustrated.

Smoothed returns generate dangerous perceptions of risk if the returns are not unsmoothed and if the short-term volatility of the returns is used without adjustment to estimate longer-term risk. The estimated quarterly standard deviation of the returns of the NPI illustrated in Exhibit 15.2 is approximately 3%. Without taking autocorrelation into account, the estimated annualized standard deviation of the same series would be 6% (found by multiplying the quarterly standard deviation by the square root of the number of quarters in each year). Based on an annualized standard deviation of 6% and an expected return of perhaps 6%, a portfolio allocator might not expect a very large annual loss. However, the correlated string of quarterly losses from the fourth quarter of 2008 to the end of 2009 shows the tremendous longer-term loss potential generated from positive autocorrelation even when short-term volatility appears modest.

Smoothed returns understate not only the volatility but also the correlation of the smoothed returns to the returns of other asset classes. Exhibit 17.11 in Chapter 17 shows the correlation between smoothed and unsmoothed return series. The stated, appraisal-based NPI returns had a correlation of −0.05 to the NAREIT Index and a zero correlation to the Russell 3000 U.S. stock index. After unsmoothing the NPI returns, the correlation between the NPI and the NAREIT Index rose to 0.37, while the correlation to the Russell 3000 rose to 0.16. Assets with low volatility and low correlations to other asset classes are highly diversifying and earn large weights in a mean-variance optimization process. Using a simple first-order autocorrelation unsmoothing process finds that the unsmoothed NPI returns have three times the volatility and twice the correlation to other asset classes as do the stated, appraisal-based returns. Using unsmoothed returns in the mean-variance process leads to substantially lower weights for private real estate in the optimal portfolio.

Underestimating risk in smoothed returns results in inflated estimates of risk-adjusted returns, including the Sharpe ratio. In fact, overestimation of risk-adjusted performance due to smoothing of returns may explain the so-called real estate risk premium puzzle, which asks why private equity real estate investments seem to offer abnormally high risk-adjusted returns relative to other investments.

The lesson is clear. Autocorrelation of returns can provide misleading indications of long-term risk relative to short-term risk. Smoothing of returns can dangerously mask true risk. Unsmoothing of returns is an important method of providing estimates that better indicate true risk and facilitate more appropriate decisions regarding asset allocation.

15.5 NOISY PRICING

As we have discussed, empirical real estate values can be obtained either by evaluating transaction prices of properties that actually sell or by using appraised values produced by professional real estate appraisers. Both types of values are noisy because they are inexact estimates of true market value (i.e., they contain error relative to the true value of real estate assets). As illustrated in Geltner et al. (2014), this error falls primarily into two major types: purely random error or noise and temporal lag bias. The **purely random error or noise** arises because of the structure of the real estate market, where transactions involve negotiations between two parties and the resulting transaction price is one value from a range of prices that could have resulted from those negotiations. **Temporal lag bias** arises when transaction prices are related to past prices because of the structure of the market. For example, the reported transaction price may have been negotiated a few months before and therefore does not reflect the current market conditions.

In addition, a trade-off between these two types of errors arises when one attempts to construct an optimal value estimation method. This occurs because it is difficult to reduce the lag bias without increasing the random error, and hard to reduce the random error without increasing the lag bias.

To understand this problem of noisy pricing, we introduce the concept of **reservation price**, which is the lowest price at which a potential seller is willing to sell a property, or the highest price a potential buyer is willing to pay for a property. Exhibit 15.3 shows that the cross-sectional reservation price distribution of potential buyers and sellers overlaps for a particular population of properties and at a certain instant in time. It is in this overlapping region (i.e., the price region where prices range between values II and IV) where real estate transactions may take place, as some sellers will have reservation prices at least as low as the reservation prices of some buyers. One can argue that these reservation prices will tend to be distributed

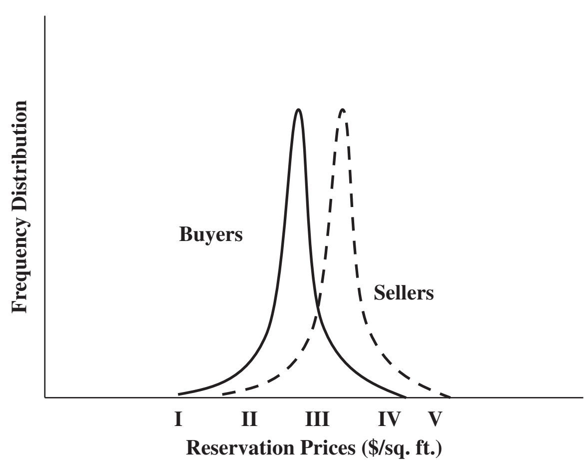


EXHIBIT 15.3 Buyers' and Sellers' Reservation Price Frequency Distributions at a Given Point in Time
Source: Based on Geltner, Miller, Clayton, and Eichholtz (2014).

around III, which is the hypothetical true market value of the population of properties at a given point in time and once they are adjusted for quality. Unfortunately, III is empirically unobservable, and all one can examine are valuations drawn from the probability distribution in III. The difference between any given price observation and the unobservable true market value is known as **transaction price noise** or **transaction price error**. Assuming that all transaction price observations occur at the same point in time, this error will be unbiased.

Real estate appraisals are also noisy because they are based on opinions (however rational and professional the appraisers may be); thus, five different appraisers may arrive at five different appraisals. The difference between a particular empirical appraised value and the unobservable true market value is known as **appraisal error**. Unlike transaction prices, the appraised-value cross-sectional dispersion may or may not be centered on the true value. This implies that appraised values may be biased. To reduce the random estimation error, more transaction data is necessary. And to obtain more transaction data, transactions must be recorded reaching further back in time. One can thus reduce the random noise, but at the expense of increasing the temporal lag, or decrease lag but increase noise. This noise versus lag trade-off has a concave form, which implies that there are diminishing returns in the process (e.g., the more transaction data that one has, the more accurate will be the inferences about real estate market values, but at a decreasing rate). More specifically, the accuracy of the estimated market value will tend to be inversely proportional to the square root of the number of transaction price data that one can examine. This is the law of statistical inference known as the square root of *n* rule. For example, if an appraiser can use four comparables instead of just two, the estimated value will have about 71% (i.e., $1/\sqrt{2}$) as much error (4/2 = 2). But if the same appraiser can use six comparables instead of just two, the estimated value will have about 58% (i.e., $1/\sqrt{3}$) as much error (6/2 = 3). Thus, one can reduce the error of the estimated value from 71% to 58% by increasing the number of comparables from four to six. However, comparable data become available only when transactions take place, and transactions occur through time. In our case, in order to use, for example, four comparables, the appraiser would have to reach, on average, twice as far back in time.

15.6 APPRAISAL-BASED REAL ESTATE INDICES

This section discusses appraisal-based indices—the first of two main types of real estate indices available to market participants. (The second type, transaction-based indices, will be discussed in the next section.) **Appraisal-based indices** are derived from the property values estimated by appraisers, which may track a particular subpopulation. In theory, all properties included in this type of index can be appraised every period, although this is not always the case.

15.6.1 Approaches to Appraisals

There are three primary means of appraising real estate: the sales comparison approach, the cost approach, and the income capitalization approach. Each method has its own value depending on the situation and type of real estate involved.

15.6.1.1 Three Approaches to Forming Appraisals These methods are most common for commercial real estate.

1. The **sales comparison approach**, in which a real estate asset is evaluated against those of comparable (substitute) properties that have recently been sold. Value adjustments may be made for characteristics such as square footage, date of sale, location, and amenities.
2. The **cost approach**, which assumes that a buyer will not pay more for a property than it would cost to build an equivalent one. A property's value can be estimated by adding the depreciated value of any improvements to the land value of the property. This approach is often suggested when valuing newer structures.
3. The **income approach**, which is similar to the discounted cash flow method used for valuing stocks and bonds. Several years of net operating income are projected for a specific property or portfolio of properties and then discounted using an appropriate discount rate. This approach is particularly useful when valuing income-producing real estate assets, such as commercial real estate.

15.6.1.2 Two Advantages of Appraisal-Based Models There are two primary advantages of appraisal-based models:

1. In general, they do not suffer from a small sample size problem (as does the repeat-sales method, which will be discussed later).
2. All properties can be appraised frequently, although this is a costly process.

15.6.1.3 Three Disadvantages of Appraisal-Based Models There are three primary disadvantages of appraisal-based models:

1. Appraisals are inherently subjective and backward-looking, thus introducing errors in prices. In the case of the NPI, not all properties are reappraised every quarter. This causes a "stale appraisal" effect (i.e., dated appraisals), which adds to the lag.
2. Appraisal-based indices are smoothed compared with actual changes in real estate market values. Thus, measures of volatility of the value of commercial real estate assets are underestimated using appraisal-based indices. Fortunately, unsmoothing techniques, such as the ones presented earlier in the chapter, help mitigate this problem. In the case of the NPI, there exists seasonality in the index due to fourth-quarter reappraisals.
3. Appraisal methods tend to rely on data from comparable properties. Therefore, the quality of the appraisal will depend critically on the quality of available data. As a result, appraisals may not be accurate in situations in which comparable properties cannot be identified (e.g., the current property is rather unique) or there is a significant time lag between the time the data has become available and the time the appraisal takes place.

15.6.2 The NCREIF Property Index, an Example of an Appraisal-Based Index

The National Council of Real Estate Investment Fiduciaries (NCREIF) Property Index (NPI) is used to illustrate appraisal-based indices due to its size and popularity. NCREIF is a U.S. not-for-profit institutional real estate investment industry association.[3] NCREIF collects data from its members, which include, for the most part, institutional real estate investment managers. NCREIF maintains a massive data set of real estate income and pricing data, and uses those data to publish the NPI and its sub-indices as well as several other indices, such as a farmland index and a timberland index.

The NPI is based on financial information from member institutional investors. Members are required to report information on their real estate holdings on a quarterly basis. The NPI started in the fourth quarter of 1977. As of the fourth quarter of 2014, it consisted of 7,062 properties (i.e., apartment, industrial, hotel, office, and retail properties) with a gross fair market value of over $409 billion. Most valuations are appraisal-based. The reason for the use of appraisals is the illiquid nature of real estate: Properties simply do not turn over frequently enough to compute short-term returns using prices from transactions performed in an arm's-length manner. NCREIF compiles the information on a value-weighted basis to publish the NPI.

As mentioned before, the NCREIF is a family of appraisal-based indices. The best known of these indices is the NCREIF Property Index, which tracks the performance of core institutional property markets in the United States on a quarterly basis. NCREIF indices are available geographically by region, division, state, and zip code. The change in value of each property in the NPI is calculated every quarter on an "as if" basis: as if the property were purchased at the beginning of the quarter at its appraised value, held for income during that quarter, and sold at the end of the quarter at its end-of-quarter appraised value. If the property was actually acquired or sold during the quarter, the transaction price is used in place of either the beginning or the ending value.

The total return on the index is calculated as the sum of an income return and a capital value return. The income portion of the total return of each property is a fraction, with net operating income (NOI) in the numerator and an estimate of the property value in the denominator. The estimate of the property value is based on the beginning-of-period appraised value, with adjustments for any capital improvements, any partial sales, and reinvestment of NOI.

The numerator of the capital value return is the change in the estimated value of the property from the beginning of the quarter to the end of the quarter, adjusted for capital improvements and partial sales such that increases in the value due to further investments are not included as profits, and declines in value due to partial sales are not deducted as losses. The denominator of the capital value return is the same as the denominator for the income portion.

The NPI is calculated on an unleveraged basis, as if the property being included in the index were purchased with 100% equity and no debt. As a result, the returns are less volatile, and there are no interest charges deducted. The returns to the NPI are calculated on a before-tax basis and therefore do not include income tax expense. The returns are calculated for each individual property and are then value-weighted in the index calculation.

As mentioned, the turnover of most real estate properties is infrequent (every six or seven years, on average), so the NPI is based primarily on appraised values rather than market transactions. The members of NCREIF report the value of their properties every quarter.

15.6.3 Two Popular Methods of Appraisals

Appraisals are professional opinions of value and are commonly used to estimate the market value of a real estate property. For institutional investors, appraisals are typically performed once a year on real estate properties. Appraisals are generally based on one of two methods (both of which are more popular than the cost approach, previously mentioned).

The first is the comparable sales approach, introduced in the previous section. In this approach, the real estate appraiser looks at sales of similar properties in the same geographic region (if not city) as the property being appraised. These actual sale prices give the appraiser an estimate of the cost (i.e., price) per square foot of similar real estate properties. The appraiser then adjusts this cost per square foot for the unique characteristics of the property being appraised: better parking or access, better location, newer lobby, longer-term tenants, and so on. This process has the advantage of being based on actual sales transactions. However, the accuracy of the process is lower when there is a lack of frequency of property sales, and, because every property is unique, it is hard to adjust a square-foot calculation value from one property to form the value of another accurately.

The second method is the income approach or discounted cash flow analysis method, also introduced in the previous section. This has become the more accepted practice by real estate appraisers for commercial properties. In this valuation approach, the appraiser estimates the cash flows from a property and discounts them to form a present value to the property. This approach has the advantage of valuing the unique characteristics of the property being appraised. However, it is subject to forecasting errors of occupancy, lease growth rates, expenses, holding period for the property, terminal value of the property, inflation estimates, and the like.

15.6.4 Illustration of the Income Approach

To illustrate the mechanics of the income approach (discounted cash flow model) using an example, suppose a German-based real estate company needs to estimate the value of an office building that it just purchased in Munich. The potential gross income and operating expenses of the first year of operations have been estimated at €10 million and €3.8 million, respectively. For simplicity, assume that these amounts are received at the end of the year. A 10% vacancy loss rate is being assumed for this investment for the first two years, decreasing to 8% for the subsequent two years. Assume that the real estate company expects to maintain the office building for four years, and that rents and operating expenses are estimated to increase by 5% per year. The office building is projected to be sold in four years, providing estimated net sales proceeds of €75 million at that time. Using discounted cash flow analysis, appraise the value of this office building. The real estate company is

EXHIBIT 15.4 Projection of Net Operating Income over Next Four Years

	Year 1	Year 2	Year 3	Year 4
Potential gross income	€10,000,000	€10,500,000	€11,025,000	€11,576,250
Vacancy loss	€ 1,000,000	€ 1,050,000	€ 882,000	€ 926,100
Effective gross income	€ 9,000,000	€ 9,450,000	€10,143,000	€10,650,150
Operating expenses	€ 3,800,000	€ 3,990,000	€ 4,189,500	€ 4,398,975
Net operating income	**€ 5,200,000**	**€ 5,460,000**	**€ 5,953,500**	**€ 6,251,175**

using a required rate of return of 6% for comparable investments. Ignore taxes for simplicity.

First, project the net operating income for the next four years (see Exhibit 15.4).

The present value of the four years of NOI and the net sales proceeds to be received in four years, discounted at 6%, is €79,122,255. This is the appraised value of the office building using the income approach.

Although the NPI is a quarterly index, NCREIF properties are not formally appraised every quarter. Most properties are formally valued at least once per year, but many are appraised only every two or even three years. Appraisals cost money; therefore, there is a trade-off between the benefits of having frequent property valuations and the costs of those valuations as a drain on portfolio performance. In fact, many institutional real estate investors value their portfolio properties only when they believe there is a substantial change in value based on new leases, changing economic conditions, or the sale of a similar property close to the portfolio property.

Even when properties have been recently appraised, it is possible that the appraisal process will be driven by old information, such as previous transactions on comparable properties, or by delays in the willingness of appraisers to adopt new valuation standards brought on by changes in market conditions, such as capitalization (cap) rates. Thus, even recent appraisals can cause smoothing due to delays in fully reflecting changes in true value. Finally, note that the NPI is published quarterly but that quarter-end values are published with a time lag. Thus, even ignoring appraisal-based smoothing, a major decline in asset prices that occurs in October would not be reflected in quarterly index figures until the December 31 appraisal, and the December 31 value would not be published until almost a month later. In contrast, market indices, such as real estate indices based on REIT market prices, are continuously updated. NCREIF has begun producing the Transaction Based Index (TBI). The TBI is a hedonic index (discussed in the next section) that uses transaction data from the NCREIF database. The index relies on a hedonic methodology based on analysis of price per square foot (i.e., property sale price divided by number of square feet in the property).

Finally, the Investment Property Databank (IPD) Index is another example of an appraisal-based index. It tracks retail, office, and industrial properties in the United Kingdom and includes data on actual property transactions from property companies and institutional investors. The IPD Index is available monthly and annually.

15.7 TRANSACTION-BASED INDICES (REPEAT-SALES AND HEDONIC)

Transaction-based real estate indices are estimated based on actual contemporaneous property sales of sample properties that trade in each period. They can be a good basis for real estate derivatives, granted that:

1. They are based on enough data and rigorous econometric methods.
2. Any differences in properties trading in different periods are controlled for.
3. Statistical noise is minimized.

On the other hand, the main problem when using transaction-based indices is that each individual property is unique and thus the problem of finding the right comparables arises when the prices of different properties are compared at the same point in time. Furthermore, each individual property is sold infrequently and erratically in time.

The two main methods used to estimate transaction-based price indices are the repeat-sales method (RSM) and the hedonic pricing method (HPM).

15.7.1 The Repeat-Sales Method

A repeat sale occurs when a specific property is sold at least two times during the sample period used. The periodic returns are then estimated from only the percentage changes in the valuation observations across time within the same properties. Thus, differences across properties are controlled for by using only price-change information from assets that have not changed. The **repeat-sales method (RSM)** regresses the percentage price changes observed in properties onto a sequence of time-dummy variables.

We now explain how a repeat-sales method can be used to construct an index based on prices estimated according to the RSM. The following example is adapted from Geltner et al. (2014). Assume two investors and two periods of time. Suppose that the first investor purchased his property for \$100,000 at the beginning of the first period and sold it for \$115,500 at the end of the second period; that is, the property was held for two full periods. Suppose the second investor purchased her property at the beginning of the second period for \$200,000 and sold it at the end of that period for \$210,000. Given this information, the time-dummy variable for the first time period would be 1 for the first investor and zero for the second investor. The time-dummy variable for the second time period would be 1 for both investors, since they both held their investments during that period. Exhibit 15.5 presents the corresponding regression model data.

The time-dummy coefficients for the two periods are then estimated by running a regression in which the estimated coefficients will correspond to the annual return on the investments represented by the two properties. In this simple example, the regression yields a 10% return as the coefficient of the dummy variable representing the first period, and a 5% return as the coefficient of the dummy variable representing the second period. The dependent variables are transformed using the log function before the regression is estimated; that is, logs of total returns, 1.155 and 1.050, are used as dependent variables.

EXHIBIT 15.5 Setup of the Regression Model Data

	Time-Dummy Variables	
Dependent Variable	**Time Period 1**	**Time Period 2**
\$115,500/\$100,000 = 1.155	1	1
\$210,000/\$200,000 = 1.05	0	1

In this example,

- For the first investor: \$100,000 × 1.10 × 1.05 = \$115,500
- For the second investor: \$200,000 × 1.05 = \$210,000

As long as data contains at least one sale per period, the repeat-sales model can find returns for each period. Statistical techniques such as ordinary least squares (OLS), weighted least squares (WLS), and time-weighted dummies can optimize the resulting estimates.

The CoStar Commercial Repeat-Sale Index (CCRSI) is an example of a repeat-sales index that measures returns for commercial real estate in the United States. It is based on 1,672 repeat sales in December 2014 and more than 130,000 repeat sales since 1996. The CCRSI family of indices report monthly and quarterly returns on 30 sub-indices in the CoStar index family. The sub-indices include breakdowns by region of the country (Northeast, South, Midwest, and West); by property sector (office, industrial, retail, multifamily, hospitality, and land); by transaction size and quality (general commercial and investment grade); and by market size (composite index of the prime market areas in the country).

15.7.1.1 Two Advantages of the Repeat-Sales Method There are two primary advantages of the repeat-sales method:

1. The RSM calculates changes in home prices based on sales of the same property. Therefore, this method avoids the problem of trying to explain price differences in properties with varying characteristics. Thus, the repeat-sales method does not require detailed information on property characteristics.
2. The RSM is relatively robust to specification error. Typically, this type of error arises when the wrong regression model is selected to explain variations in the dependent variables. Specification error causes the independent variable and the error term of a regression equation to become correlated.

15.7.1.2 Three Disadvantages of the Repeat-Sales Method There are three primary disadvantages of the repeat-sales method:

1. The main disadvantage of the RSM is that the sample of properties from which the price changes and the resulting index are calculated represents a very small portion of all the properties that are transacted during a given period of time. This data scarcity problem can be severe when trying to apply this method during

a short period of time. This disadvantage can be important because properties that were sold only once during a certain period of time (and are therefore not considered when using the RSM) are also important signs of real estate market movements. As a result, the RSM may suffer from a sample selection bias.
2. The RSM assumes that the property being traded does not experience changes, when in reality all properties depreciate and age, and some are renovated.
3. When the RSM is updated, it generates backward adjustments in the historical series of returns, and thus previously estimated returns for certain years in the past may change. This problem occurs because a property recording a new "second sale" links back to the earlier "first sale" in estimation history.

15.7.2 The Hedonic Pricing Method

The price of every real estate asset is determined by the overall supply and demand conditions in the local real estate market, as well as by the different set of attributes that the asset commands. The **hedonic pricing method (HPM)** assumes that each of these attributes has its own market, and that the price of each attribute is determined by its demand and supply. Examples of these attributes are the number of rooms, the size of the lot, the number of bathrooms, and so on.

Real estate assets are heterogeneous goods, which renders them virtually unrepeatable and unique. We can observe the market price of a property, but we cannot disentangle the marginal prices of the attributes that make it up. It is thus required to determine the implicit price or contribution of each of these attributes (i.e., the hedonic price) within the observed market price.

For the HPM, the market price of a real estate asset is an aggregate of the individual prices of all the attributes (which may be heterogeneous) that make up the property. The attributes of a property may be internal (location, size, number of rooms, age) or external (accessibility to schools, crime rate in the neighborhood, level of air pollution). The HPM is based on actual market prices and is relatively inexpensive to apply when data are readily available.

The concept of a hedonic price index applied to real estate transactions involves using observed real estate transactions of some properties to estimate the prices of all properties, including those that did not transact. Hedonic price indices infer the prices of real estate properties that have not recently traded by directly modeling the heterogeneity of real estate properties.

15.7.2.1 Three Steps to Calculating a Hedonic Price Index The three steps of calculating a hedonic price index using transaction prices are:

1. Model the value of real estate properties as being a function of specified characteristics of the properties.
2. Use a sample of prices observed from recent transactions to fit the parameters of the real estate valuation model; that is, determine the implicit price of each attribute.
3. Use the estimated valuation parameters to estimate the values of the properties that are within the index but did not transact.

As a very simplified example, a model of office properties may be valued using a model with two variables: size (square feet of space) and quality (Class 1, 2, or 3). Data on recent office property transactions are recorded with a sale price, size, and

EXHIBIT 15.6 Hypothetical Prices, Sizes, and Qualities of Six Office Properties

	Sale Price	Size (Sq. Ft.)	Quality
Office A	$110,000	840	3
Office B	$145,000	1,155	2
Office C	$152,000	1,035	1
Office D	$133,000	970	2
Office E	$205,000	1,250	1
Office F	$ 95,000	760	3

quality class for each. The parameters for the size and quality variables are estimated based on an econometric model, with price per square foot as the dependent variable, and size and quality as the independent variables. The estimated parameters from the model are then used to infer the prices of all properties. These prices are then used to calculate prices of other properties and the index representing the overall market.

We now illustrate how the hedonic regression model can be used to estimate property prices. Let us suppose that six offices (A through F) of different sizes (in square feet) and quality types (from 1 to 3, where 1 is the highest) have recently been sold at the prices shown in Exhibit 15.6. How can one use this information to predict the price of a seventh office, G?

First, estimate a regression equation using the OLS, where the natural logarithm of price is the dependent variable, and size and quality are the independent variables. Then calculate the natural logarithm (ln) of price, and use it as the dependent variable because it provides a better estimation when estimating hedonic regressions. The estimated regression equation is as follows (note that t-statistics are in parentheses):

$$\begin{aligned}\ln(\text{Price}) = 11.13151 + {} & \underset{(3.06)}{0.000918} \times \text{Size} - \underset{(-1.86)}{0.1159316} \times \text{Quality}\end{aligned}$$

The R^2 of the regression is 0.95, and the variables' size and quality have the expected sign and are statistically significant at 5% and 10% levels, respectively. Let us now suppose that we need to estimate the price of a new office, G, which is of quality type 2 and has 1,020 square feet. Using the estimated equation, we get:

$$\ln(\text{Price}) = 11.13151 + 0.000918 \times 1,020 - 0.1159316 \times 2$$
$$\ln(\text{Price}) = 11.8360068$$

Calculating $e^{\ln(\text{price})}$, we get the estimated price of office G:

$$e^{14.124391} = \$138,136$$

15.7.2.2 Five Advantages of the Hedonic Pricing Model There are five primary advantages of the hedonic pricing model:

1. The HPM uses all observations and not just repeat sales. This advantage is particularly useful when the length of the sample is short.

2. The HPM is versatile, as hedonic indices can be adapted to take into consideration the several probable interactions between environmental quality and marketed goods.
3. The HPM avoids backward adjustments of historical returns when an index is reestimated with "second sale" transactions data.
4. The HPM allows analysis of a property's value attributes.
5. The HPM can also be used to estimate the marginal contribution to property equilibrium prices of certain attributes in the real estate market that are either unobservable or hard to measure. For example, the HPM has been used to estimate the effect of the following on housing prices: the presence of air pollution, scenic views versus industrial views, the perceived risk of only partially compensated expropriations, and the existence of open space and protected areas.

15.7.2.3 Six Disadvantages of the Hedonic Pricing Model There are six primary disadvantages of the hedonic pricing model:

1. The HPM requires a large amount of data on several hedonic (both internal and external) variables. The collection of this data can be costly and time-consuming.
2. As in the case of the repeat-sales model, the HPM may also suffer from sample selection bias. This arises when the properties that are sold are not representative of the universe of properties. For example, a sample selection bias will arise if, in a normal market, the owners of properties with rising values tend to sell their properties, while the owners of properties with falling values tend not to sell their properties.
3. The HPM is exposed to specification error. That is, all attributes of property may not have been entered into the model.
4. The HPM assumes that everyone knows beforehand the potential positive and negative externalities that arise when purchasing a specific property. For example, the model assumes that buyers already know about the noise level in a locality situated near a factory when in reality they may not.
5. The HPM assumes that, for a given income level, buyers will be able to select the combination of attributes that they prefer.
6. An econometric problem (multicollinearity) may arise if, for example, larger properties are available only in the safer areas of cities, and smaller properties are available only in crime-driven areas. In such cases, it would be difficult to exactly disentangle the effects of crime and property size on prices.

15.7.3 Contrasting Repeat-Sales and Hedonic Price Indices

The primary difference between hedonic price indices and repeat-sales indices is that hedonic price indices include all properties that transacted at least once, whereas repeat-sales indices include data from only those properties that transacted multiple times. A hedonic price index uses observed price levels on some real estate to infer price levels on all real estate. A repeat-sales index uses observed changes in prices on some real estate to infer changes in prices on all real estate. The reliability of both approaches depends on the accuracy of the model, including the effects of property differences and timing differences.

The MIT-developed NCREIF TBI is an example of a family of real estate indices that are estimated using the hedonic pricing model but have similarities to the repeat-sales method indices. The index is based on the NCREIF Index database and thus is an institutional real estate index. The TBI regresses transaction prices from properties sold from the NCREIF Property Index database onto their recent appraised values and time-dummy variables. The index combines the NCREIF Property Index (which is appraisal-based) with the regression time-dummy coefficients and thus reflects the differences between appraised values and transaction prices. The TBI was launched by the MIT Center for Real Estate (MIT/CRE) in February 2006 and published quarterly. In 2011, NCREIF took over production and publication of the index (which is now known as NTBI) and slightly modified its methodology.

15.7.4 Sample Biases in Transaction-Based Indices

Transaction-based real estate indices suffer from sample selection bias to the extent that the properties that are represented in the index differ from the aggregated mix of all such assets. For example, in transaction-based indices, a bias arises to the extent that properties transacted during a particular period and used to calculate an index may not be representative of the entire universe of properties. A bias toward a particular type of property, location of property, size of property, and so forth, would interfere with the representativeness of the sample. Although the effects of sample selection bias also influence residential property indices, they are likely to be especially severe for commercial properties because the universe of commercial properties is typically small and the number of transactions in any particular time period is even smaller. The effects of randomness on the attributes of a sample tend to be larger when the sample is small.

There is another reason to be concerned about sample selection bias other than the idea that the sample of properties with transactions will deviate from the universe of all properties due to randomness. As Haurin (2005) points out,

> *In a normal market, the real values (i.e., deflated values) of some properties will rise while others may decline. If the owners of properties with falling values tend to choose not to sell their properties, while owners of properties with rising values tend to choose to sell (or vice versa), then the sample of transacted properties is clearly not random and is biased towards a particular price outcome. It is also plausible that the choices of whether to sell properties with rising and falling values change over the real estate cycle and thus the nature of the sample selection bias will change over time. This changing bias results in an estimated transaction-based price index that differs from a theoretical price index that would track market values of the stock of all properties.*

Thus, there may be systematic reasons to believe that the sample of properties that have transacted will have substantially different characteristics from the universe of properties and that, more importantly, the differences in characteristics will be related to recent price changes. The result is that the sample will reflect price changes unrepresentative of the universe.

15.8 DESCRIPTION OF MAJOR REAL ESTATE INDICES

This section offers a brief description of the main real estate indices available, divided into various categories. As indicated in Exhibit 15.7, most of the major real estate indices in the world are based on market prices of real estate securities. Real estate indices can differ as to whether they are indices of values (prices) or total return (including income). Those indices based on total return can also differ in terms of whether the returns are gross or net of fees.

15.8.1 Housing or Residential Real Estate Properties Indices

Institutional investors have limited investment interest in residential real estate equity other than multifamily properties (i.e., apartment buildings, which are usually classified as commercial). However, residential real estate prices assist in the analysis of residential mortgages and provide economic data useful in general economic and real estate analysis. In addition, derivatives can be used to provide institutional investment in residential real estate equity. Accordingly, residential real estate indices are included in Exhibit 15.7.

A popular but primitive type of index of residential property prices is derived from the reporting of mean or median housing prices over various time periods. The major problem with these indices as an indication of real estate value changes is that they are affected by shifts in the quality and size of the properties that happen to transact within each time period. Further, observed prices may reflect changes in the condition of the properties, especially substantial improvements.

The values of U.S. homes or residential real estate properties are tracked by the Standard & Poor's (S&P)/Case-Shiller Home Price Indices, which consist of metropolitan regional indices, composite sub-indices, and a national index. The indices are constructed using the repeat-sales method, which requires observation of the sale prices of specific single-family homes that are sold at least twice within the observation period. The new sale price is combined with the previous sale price to form a sales pair. The price differences in the sales pairs within a particular region are measured and used to infer changes in the levels of the index for that region. This family of indices has been calculated monthly since January 1987. Options and futures based on the S&P/Case-Shiller indices are traded on the Chicago Mercantile Exchange.

15.8.2 Farmland and Timberland Indices

The NCREIF Farmland Index (not shown in Exhibit 15.7) is a quarterly appraisal-based index that measures the investment performance of a large pool of individual agricultural properties acquired in the private market solely for investment purposes. Only income-generating agricultural properties are included in the index. According to NCREIF, all properties in the Farmland Index have been acquired, at least in part, on behalf of tax-exempt institutional investors, the great majority being pension funds. As such, returns reflect properties held in a fiduciary environment.

EXHIBIT 15.7 Summary of Popular Real Estate Indices

Index Name	Region	Methodology	Type of Property	Source of Values
S&P/ASX 200 A-REIT Index	Australia	Market	Commercial	REIT
Bloomberg Canadian REIT	Canada	Market	Commercial	Closed-end trust
Shanghai Stock Exchange Property Index	China	Market	Commercial	Stock
EURO STOXX Real Estate (Price) EUR	EU	Market	Commercial	Stock
Bloomberg Europe 500 Real Estate Index	Europe	Market	Commercial	Stock
IPD Germany Annual Property Index	Germany	Repeat sales	Commercial	Private
Hong Kong Hang Seng Properties Index	Hong Kong	Market	Commercial	Stock
Tokyo Stock Exchange REIT Index	Japan	Market	Commercial	REIT
FTSE ST Real Estate Index	Singapore	Market	Commercial	Stock
SWX IAZI Real Estate Performance Index	Switzerland	Hedonic	Commercial	Private
SWX IAZI Private Real Estate Price Index	Switzerland	Hedonic	Residential	Private
FTSE 350 Real Estate Supersector Index	UK	Market	Commercial	Stock
FTSE UK Commercial Property Index	UK	Appraisal	Commercial	Private
IPD UK Annual All Property Index	UK	Appraisal	Commercial	Private
FTSE NAREIT All Equity REITs Index	U.S.	Market	Commercial	REIT
NCREIF Property Index (NPI)	U.S.	Appraisal	Commercial	Private
NCREIF Transaction Based Index (TBI)	U.S.	Hedonic	Commercial	Private
S&P/Case-Shiller U.S. National Home Price Index	U.S.	Repeat sales	Residential	Private
CoStar Commercial Repeat-Sale Indices	U.S.	Repeat sales	Commercial	Private
Green Street Advisors Commercial Property Price Index	U.S.	Mixed	Commercial	Private*

*Reflects valuations based on professional analysis of property owned by listed equity REITs.

The NCREIF Timberland Index (not shown in Exhibit 15.7) is a quarterly index that measures the investment performance of institutional timberland investments. To qualify for the index, a property must be held in a fiduciary environment and marked-to-market at least once per year. The lack of quarterly appraisals for many properties in this index makes the annual return series more reflective of changes in the market than the quarterly series.

15.8.3 Market-Traded Real Estate Vehicles

A final major valuation method for real estate indices is based on tracking the market prices of investment vehicles with underlying real estate assets. Many countries, including the United States, Australia, France, and the United Kingdom, have publicly traded real estate investment trusts (REITs). REITs typically contain leverage and differ in underlying real estate investments (e.g., mortgage versus equity, office versus retail). Nevertheless, many REITs are traded with regular and substantial volume. Thus, indices based on market prices of real estate–related securities are becoming increasingly important in the monitoring of real estate market conditions.

There are numerous indices based on real estate securities. In the United States, the FTSE NAREIT US Real Estate Index Series consists of a family of REIT performance indices that covers the different sectors of the U.S. commercial real estate space. Constituents of the FTSE NAREIT Composite Index are classified as either equity REITs (which predominantly own equity real estate in the form of buildings) or mortgage REITs (which predominantly own real estate debt). Equity REITs can be further subdivided by property sector: industrial/office, retail, residential (apartments and manufactured homes), diversified (specialized), lodging/resorts, health care, self-storage, and timber. The industrial/office and retail sectors are further broken down into three subsectors each.

Other widely used public real estate equity indices (not shown in Exhibit 15.7) are the S&P U.S. REIT Composite Index, the Dow Jones Wilshire Real Estate Investment Trust Index (DJW REIT), the Dow Jones Wilshire Real Estate Securities Index (RESI), and the MSCI US REIT Index. The following are some of the REIT indices available around the world: in the United Kingdom, the FTSE EPRA/NAREIT UK REIT Index; in Australia, the S&P/ASX 200 A-REIT Index; and in Japan, the Tokyo Stock Exchange REIT Index (TSEREIT).

15.8.4 Real Estate Debt or Mortgage Indices

In the United States, secondary trading of pools of residential mortgages is a large and well-developed market, with substantial and prompt revelation of values. Prices are revealed through the pricing of a number of investment vehicles and indices, including pass-through pools, structured products such as residential mortgage-backed securities (RMBS), and numerous mortgage-backed security (MBS) indices. Furthermore, the FTSE NAREIT Mortgage REIT Index provides an indication of mortgage values through the public valuation of mortgage REITs.

The pricing and performance of private commercial mortgages in the United States can be measured using a wide offering of indices, especially those that track the value of public commercial real estate debt structured as commercial mortgage-backed securities (CMBS).

15.9 REAL ESTATE INDICES PERFORMANCE

With such a wide variety of real estate indices available, it is important for investors to understand the performance, risks, and return drivers of the indices.

15.9.1 Real Estate Indices Performance

Exhibit 15.8 shows the evolution of Moody's RCA CPPI, the NPI, the NCREIF TBI, and NAREIT between 2000 and 2014. The Equity NAREIT, which will be analyzed in more detail in a future section, provided the highest return, although it was also more volatile than Moody's RCA CPPI and the NPI. This is not surprising, as the latter two indices are appraisal-based and suffer from data smoothing, as discussed earlier in the chapter. It is also salient to notice the higher volatility exhibited by the NCREIF TBI (a transaction-based index) compared to the NPI (an appraisal-based index). It can also be observed that toward the end of 2008, at the beginning of the global financial and real estate crisis, the NPI peaked and then declined before the NCREIF TBI did. Once again, these last two findings are consistent with appraisal-based indices being smoothed.

15.9.2 Mortgage REITs Performance

Exhibits 15.9 and 15.10 show that between January 2000 and December 2014, an investment in mortgage REITs increased by 11.1% per year. Mortgage REITs

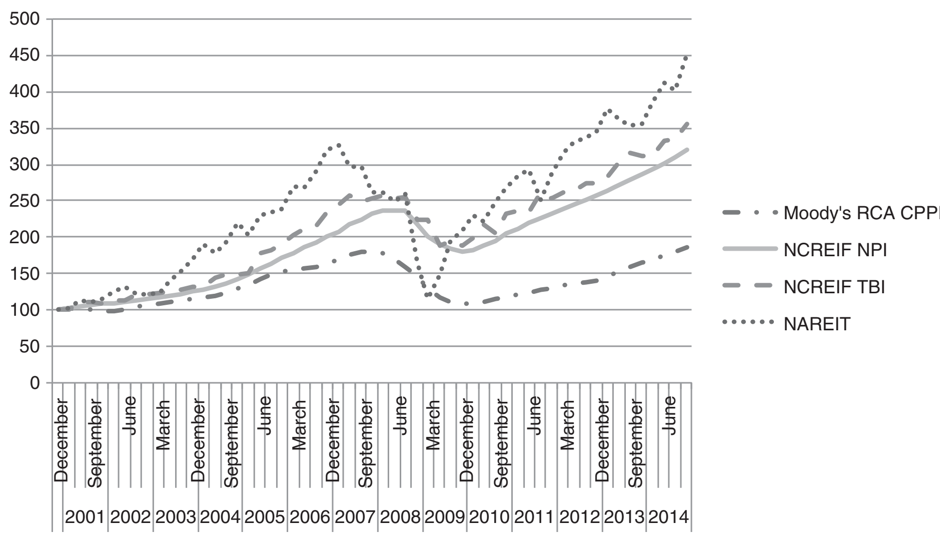


EXHIBIT 15.8 Evolution of Moody's RCA CPPI, NPI, NCREIF TBI, and NAREIT, 2000–14 (December 2000 = 100)

Note: Moody's RCA CPPI is the (U.S.) Property Index, NPI is the NCREIF (U.S.) Property Index, NCREIF TBI is the (U.S.) Transaction Based Index, and NAREIT is the All (U.S.) REITs Total Return Index.

Sources: www.rcanalytics.com, www.ncreif.org/property-index-returns.aspx, and www.reit.com/investing/index-data/monthly-index-values-returns.

EXHIBIT 15.9 Statistical Summary of Returns

Index (Jan. 2000–Dec. 2014)	Mortgage REITs	World Equities	Global Bonds	U.S. High Yield	Commodities
Annualized arithmetic mean	11.1%**	4.4%**	5.7%**	7.7%**	3.8%**
Annualized standard deviation	20.4%	15.8%	5.9%	10.0%	23.3%
Annualized semivariance	2.3	1.4	0.2	0.5	3.0
Skewness	−1.3**	−0.7**	0.1	−1.0**	−0.5**
Kurtosis	3.7**	1.5**	0.6*	7.7**	1.3**
Sharpe ratio	0.44	0.14	0.60	0.56	0.07
Sortino ratio	0.058	0.018	0.080	0.082	0.010
Annualized geometric mean	9.0%	3.1%	5.5%	7.2%	1.1%
Annualized standard deviation (autocorrelation adjusted)	22.8%	18.3%	6.2%	13.3%	27.9%
Maximum	14.2%	11.2%	6.6%	12.1%	19.7%
Minimum	−24.1%	−19.0%	−3.9%	−15.9%	−28.2%
Autocorrelation	12.3%*	16.0%**	6.1%	30.7%**	19.4%**
Max drawdown	69.1%	−54.0%	−9.4%	33.3%	−68.4%

*Significant at 90% confidence.
**Significant at 95% confidence.
Note: Mortgage REIT returns were measured from FTSE's NAREIT Mortgage REITs Index, world equities reflect the total returns of MSCI's World Equities Index, global bond returns were measured using JPM's Total Aggregate Global Index Returns, U.S. high yield is based on returns to Barclays U.S. High Yield, and commodities' returns reflect the returns to the S&P's GSCI.
Source: Bloomberg.

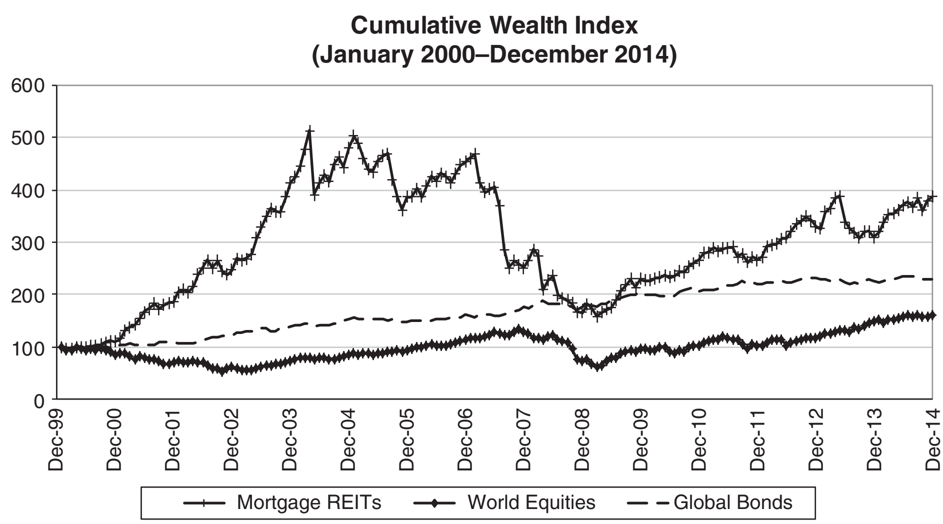


EXHIBIT 15.10 Cumulative Wealth: Mortgage REITs, World Equities, and Global Bonds, 2000–14
Note: Mortgage REIT returns were measured from FTSE's NAREIT Mortgage REITs Index, world equities reflect the total returns of MSCI's World Equities Index, and global bond returns were measured using JPM's Total Aggregate Global Index Returns.
Source: Bloomberg.

EXHIBIT 15.11 Betas and Correlations between Mortgage REITs and Four Asset Classes and Two Risk Variables, 2000–14

Multivariate Betas	World Equities	Global Bonds	U.S. High Yield	Commodities	Annualized Estimated α	R^2
Mortgage REITs	0.34**	0.58**	0.39**	−0.14**	4.20%	0.20**

Univariate Betas	World Equities	Global Bonds	U.S. High Yield	Commodities	%Δ Credit Spread	%Δ VIX
Mortgage REITs	0.48**	0.84**	0.74**	0.02	−0.04	−0.13**

Correlations	World Equities	Global Bonds	U.S. High Yield	Commodities	%Δ Credit Spread	%Δ VIX
Mortgage REITs	0.37	0.24	0.36	0.03	−0.07	−0.43

*Significant at 90% confidence.
**Significant at 95% confidence.
Note: Mortgage REIT returns were measured from FTSE's NAREIT Mortgage REITs Index, world equities reflect the total returns of MSCI's World Equities Index, global bond returns were measured using JPM's Total Aggregate Global Index Returns, U.S. high yield is based on returns to Barclays U.S. High Yield, commodities' returns reflect the returns to the S&P's GSCI, percentage change in credit spread is calculated as the difference between yields to maturity on Moody's Corporate BAA and Moody's Corporate AAA, and percentage change in VIX is the percentage change in the volatility index.
Source: Bloomberg.

provided higher returns than global stocks or bonds but at a higher risk during that period. The Sharpe and Sortino ratios for mortgage REITs were higher than those of world equities and commodities but lower than those of global bonds and U.S. high yield bonds (mainly because of the very high standard deviation exhibited by mortgage REITs). Mortgage REITs also exhibited the largest maximum drawdown (69.1%) of the five asset classes included in the table (mortgage REITs, world equities, global bonds, U.S. high yield bonds, and commodities). This maximum drawdown occurred during the global real estate and financial crises of 2008–9. In sum, mortgage REITs provided the highest returns of the five asset classes considered here but at the expense of a relatively high risk, as measured by standard deviation, semivariance, maximum drawdown, and kurtosis.

Exhibit 15.11 shows that each multivariate beta of mortgage REIT returns with respect to the returns of the other four asset classes is statistically significant at the 5% level (except for commodities, whose beta is slightly negative). Mortgage REITs provided a respectable annualized alpha of 4.20% between 2000 and 2014. Furthermore, the correlation between mortgage REIT returns and the returns of the first four asset classes is fairly low, a finding that suggests that including mortgage REITs in a portfolio may provide diversification benefits. As expected, the correlations between mortgage REIT returns and the percentage variation in credit spreads and the percentage variation in the VIX (two measures of risk) are negative.

EXHIBIT 15.12 Statistical Summary of Returns

Index (Jan. 2000–Dec. 2014)	Equity REITs	World Equities	Global Bonds	U.S. High Yield	Commodities
Annualized arithmetic mean	14.6%**	4.4%**	5.7%**	7.7%**	3.8%**
Annualized standard deviation	22.2%	15.8%	5.9%	10.0%	23.3%
Annualized semivariance	2.3	1.4	0.2	0.5	3.0
Skewness	−0.9**	−0.7**	0.1	−1.0**	−0.5**
Kurtosis	7.1**	1.5**	0.6*	7.7**	1.3**
Sharpe ratio	0.56	0.14	0.60	0.56	0.07
Sortino ratio	0.081	0.018	0.080	0.082	0.010
Annualized geometric mean	12.1%	3.1%	5.5%	7.2%	1.1%
Annualized standard deviation (autocorrelation adjusted)	23.1%	18.3%	6.2%	13.3%	27.9%
Maximum	31.0%	11.2%	6.6%	12.1%	19.7%
Minimum	−31.7%	−19.0%	−3.9%	−15.9%	−28.2%
Autocorrelation	4.2%	16.0%**	6.1%	30.7%**	19.4%**
Max drawdown	−68.3%	−54.0%	−9.4%	−33.3%	−69.4%

*Significant at 90% confidence.
**Significant at 95% confidence.
Note: Equity REIT returns were measured from FTSE's NAREIT Equity REITs Index, world equities reflect the total returns of MSCI's World Equities Index, global bond returns were measured using JPM's Total Aggregate Global Index Returns, U.S. high yield is based on returns to Barclays U.S. High Yield, and commodities' returns reflect the returns to the S&P's GSCI.
Source: Bloomberg.

15.9.3 Equity REIT Performance

Exhibits 15.12 and 15.13 show that between January 2000 and December 2014 an investment in equity REITs sextupled its value in dollar terms, in spite of the large drawdown that occurred in the midst of the global real estate and financial crises of 2008–9 (–68.3%). Equity REITs provided a substantially higher return than did world equities (14.6% versus 4.4%) or global bonds but at a higher risk during that period. The Sharpe and Sortino ratios for equity REITs were higher than those of world equities and commodities, and about the same as those of global bonds and U.S. high yield (mainly because of the very high standard deviation exhibited by equity REITs, which compensated the high returns exhibited by this asset class). Equity REITs and commodities recorded the largest maximum drawdowns of the five asset classes included in the table (equity REITs, global stocks, global bonds, U.S. high yield, and commodities). In sum, and similar to the case of mortgage REITs discussed before, equity REITs provided the highest returns of the five asset classes considered here, but at the expense of a relatively high risk, as measured by standard deviation, semivariance, maximum drawdown, and kurtosis.

Exhibit 15.14 shows that each multivariate beta of equity REIT returns with respect to the returns of the other four asset classes is statistically significant at the 5% level (except for commodities, whose beta is slightly negative, although it is not

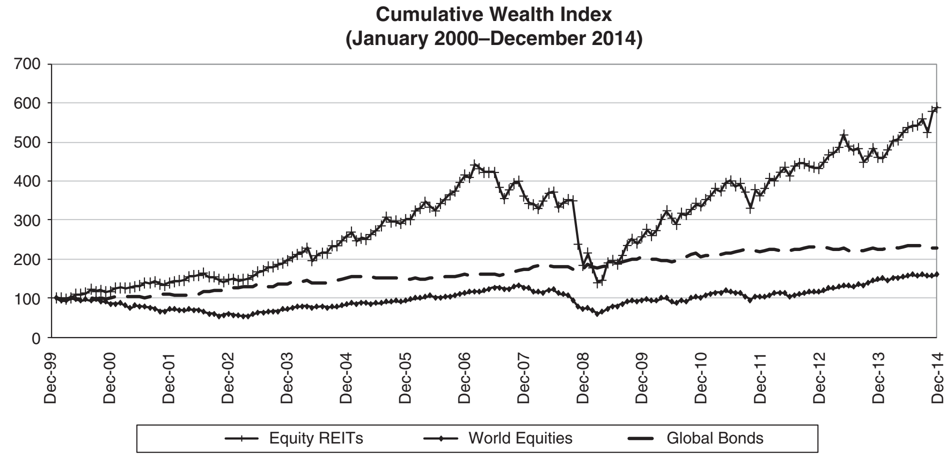


EXHIBIT 15.13 Cumulative Wealth: Equity REITs, World Equities, and Global Bonds, 2000–14
Note: Equity REITs returns were measured from FTSE's NAREIT Equity REITs Index, world equities reflect the total returns of MSCI's World Equities Index, and global bond returns were measured using JPM's Total Aggregate Global Index Returns.
Source: Bloomberg.

EXHIBIT 15.14 Betas and Correlations between Equity REITs and Four Asset Classes and Two Risk Variables (2000–14)

Multivariate Betas	World Equities	Global Bonds	U.S. High Yield	Commodities	Annualized Estimated α	R^2
Equity REITs	0.55**	0.47**	0.79**	−0.09	5.28%	0.49**

Univariate Betas	World Equities	Global Bonds	U.S. High Yield	Commodities	%Δ Credit Spread	%Δ VIX
Equity REITs	0.89**	1.08**	1.40**	0.18**	−0.15**	−0.14**

Correlations	World Equities	Global Bonds	U.S. High Yield	Commodities	%Δ Credit Spread	%Δ VIX
Equity REITs	0.63	0.29	0.63	0.19	−0.24	−0.43

*Significant at 90% confidence.
**Significant at 95% confidence.
Note: Equity REIT returns were measured from FTSE's NAREIT Equity REITs Index, world equities reflect the total returns of MSCI's World Equities Index, global bond returns were measured using JPM's Total Aggregate Global Index Returns, U.S. High Yield is based on returns to Barclays U.S. High Yield, commodities' returns reflect the returns to the S&P's GSCI, percentage change in credit spread is calculated as the difference between yields to maturity on Moody's Corporate BAA and Moody's Corporate AAA, and percentage change in VIX is the percentage change in the volatility index.
Source: Bloomberg.

statistically significant). Equity REITs provided a high annualized alpha of 5.28% between 2000 and 2014. Also, each multivariate and univariate beta of equity REIT returns with respect to the returns of the other four asset classes is higher than the respective multivariate betas and statistically significant at the 5% level. Furthermore, the correlation between equity REIT returns and the returns of global equity and U.S. high yield is relatively high. The correlations between equity REITs and the rest of the assets or variables are lower. As expected, the correlations between equity REIT returns and the percentage variation in credit spread and the percentage variation in the VIX are negative.

15.10 CONCLUSION

This chapter has addressed real estate indices and the difficulties of working with smoothed returns. Smoothed returns arise when real estate indices are calculated using appraisal-based methodologies where lagged prices are used to calculate the current value of properties. It is dangerous to use uncorrected prices of appraised real estate in asset allocation models, as the appraisals understate the volatility of real estate prices and the correlation of real estate returns to the returns of other asset classes, which leads to larger than desired weights to real estate in the optimal portfolio. Smoothed prices can be detected through the autocorrelation coefficient, where higher degrees of autocorrelation show a greater degree of price smoothing. The unsmoothing process reduces the degree of autocorrelation in a smoothed return series, which makes the resulting returns more appropriate for use in asset allocation models and a more direct comparison to liquid real estate investments such as REITs.

Real estate indices can be calculated using appraisals or transactions. When appraisals are used, the value of a given property is determined by comparing its characteristics to similar properties that have sold in the recent past. The appraisal process introduces lags into property valuation, as the subject property is valued using past transactions. Appraised values may not closely match the eventual sale price of a building due to the idiosyncratic nature of specific buildings. Transactions-based indices have less smoothing than appraisal-based indices, as returns are based on repeat sales of single properties.

NOTES

1. To demonstrate the details, the following equation is identical to Equation 15.1 except that the relationship is expressed for the reported price of the previous time period ($P_{t-1,\ \text{reported}}$) rather than the current time period ($P_{t,\ \text{reported}}$):

$$P_{t-1,\ \text{reported}} = \alpha + \beta_0 P_{t-1,\ \text{true}} + \beta_1 P_{t-2,\ \text{true}} + \beta_2 P_{t-3,\ \text{true}} + \cdots$$

Subtracting each side of the equation from the respective sides of Equation 15.1—assuming that the parameters are constants—and rearranging the terms generates this equation:

$$\Delta P_{t,\ \text{reported}} = \beta_0 \Delta P_{t,\ \text{true}} + \beta_1 \Delta P_{t-1,\ \text{true}} + \beta_2 \Delta P_{t-2,\ \text{true}} + \cdots$$

where Δ indicates single-period price changes (e.g., $\Delta P_{t,\text{ reported}} = P_{t,\text{ reported}} - P_{t-1,\text{ reported}}$). Thus, the modeling of prices is similar to the modeling of price changes. However, returns are not equivalently modeled. For example, in order to convert each price change in the preceding equation into returns, each price change must be divided by its initial price, but the prices differ.

2. The fact that ρ is the first-order autocorrelation coefficient can be verified by solving for the correlation coefficient between $R_{t-1,\text{reported}}$ and the right-hand side of Equation 15.8 (the formula for $R_{t,\text{reported}}$). Under the realistic assumption that $R_{t,\text{true}}$ is uncorrelated with the previous reported return, $R_{t-1,\text{reported}}$, the correlation between $R_{t,\text{reported}}$ and $R_{t-1,\text{reported}}$ is easily shown to be ρ.
3. Information in this chapter regarding the NCREIF Property Index is from the "Users Guide to the NPI," www.ncreif.org/public_files/Users_Guide_to_NPI.pdf, accessed January 2012.

REFERENCES

Fisher, J. D. 2005. "U.S. Commercial Real Estate Indices: The NCREIF Property Index." BIS White Papers 21.

Gallais-Hamonno, G., and H. Nguyen-Thi-Thanh. 2007. "The Necessity to Correct Hedge Funds Returns: Empirical Evidence and Correction Method." Working Paper CEB 07-034.RS, ULB—Université Libre de Bruxelles.

Geltner, D., N. Miller, J. Clayton, and P. Eichholtz. 2014. *Commercial Real Estate Analysis and Investments*. Mason, OH: OnCourse Learning.

Haurin, D. R. 2005. "US Commercial Real Estate Indices: Transaction-Based and Constant-Liquidity Indices." BIS White Papers 21, April.

CHAPTER 16

Investment Styles, Portfolio Allocation, and Real Estate Derivatives

Chapter 14 provided numerous examples of distinctions that portfolio allocators can make when categorizing real estate. As detailed in that chapter, the primary subgroup on which real estate equity analysis is focused is commercial (income-producing) real estate. This chapter focuses on the analysis of commercial real estate by institutional investors. It also introduces the most important real estate derivatives available in the marketplace, as well as their uses by institutional investors.

Institutional access to private commercial real estate investment is typically through limited partnerships involving one or several properties. An institution purchases all of the limited partnership interests or a portion thereof (along with other institutions) to have exposure to the underlying assets as well as to enjoy the managerial services of the general partner.

Styles are a primary method by which asset allocators can provide organization and structure to their universe of available investments. For example, in public equities, the styles of growth and value are often used to classify equities. By categorizing investments, an asset allocator attempts to construct better portfolios, use more appropriate benchmarks, and perform better return attribution.

The premier approach to organizing investments within the category of private commercial real estate equity is through styles of real estate investing, which refers to the categorization of real estate managers or real estate investments based on risk and return expectations. In 2003, the National Council of Real Estate Investment Fiduciaries (NCREIF) defined styles within real estate investment. Specifically, NCREIF identified three styles that relate to the underlying assets of commercial real estate investing: core, value added, and opportunistic. These styles may be thought of as a way of classifying either real estate equity investment or real estate managers.

The primary purpose of using styles of real estate investing is to provide a framework within which an asset allocator can structure top-down investment management decisions. In other words, real estate investment styles assist an asset allocator in organizing and evaluating real estate opportunities. Real estate investment styles also facilitate improved bottom-up investment management to the extent that they allow more appropriate benchmarking and performance attribution. Furthermore, style analysis can allow asset allocators to better understand the strategy being pursued by a manager of real estate funds and to monitor style drift.

16.1 DEFINING THE THREE NCREIF REAL ESTATE STYLES

The three NCREIF styles divide real estate opportunities from least risky (core) to most risky (opportunistic), with value added in the middle. In terms of risk, core properties are most bond-like, and opportunistic properties are most equity-like. Core properties tend to offer reliable cash flows each year from rents and lease payments, whereas opportunistic properties offer potential capital appreciation and typically have little or no currently reliable income. Each of the three styles is more fully described in the following paragraphs.

Core real estate includes assets that achieve a relatively high percentage of their returns from income and are expected to have low volatility. Core properties are the most liquid, most developed, least leveraged, and most recognizable properties in a real estate portfolio. Though these properties have the greatest liquidity, they are not sold as quickly as traditional investments are. Core properties tend to be held for a long time to take full advantage of the lease and rental cash flows that they provide. The majority of their returns comes from cash flows rather than from value appreciation, and very little leverage is applied. Core properties are somewhat bond-like in the reliability of their income.

Value-added real estate includes assets that exhibit one or more of these three attributes:

1. Achieving a substantial portion of their anticipated returns from appreciation in value
2. Exhibiting moderate volatility
3. Not having the reliability of core properties

Value-added properties begin to stray from the more common and lower-risk real estate investments included in the core real estate style. The value-added real estate style includes hotels, resorts, assisted-care living facilities, low-income housing, outlet malls, hospitals, and the like. These properties tend to require a subspecialty within the real estate market to be managed well and can involve repositioning, renovation, and redevelopment of existing properties.

Relative to core properties, value-added properties are anticipated to produce less income and to rely more on property appreciation to generate total return. However, property appreciation is subject to great uncertainty, and value-added properties as a whole have experienced prolonged periods of poor realized appreciation.

Value-added properties can also include new properties that would otherwise be core properties except that they are not fully leased, such as a new apartment complex or a new shopping center. A value-added property can also be an existing property that needs a new strategy, such as a major renovation, new tenants, or a new marketing campaign. These properties tend to use more leverage and generate a total return from both capital appreciation and income.

Pennsylvania's Public School Employees' Retirement System (PSERS) identifies value-added real estate as follows:

> *Value-added real estate investing typically focuses on both income and growth appreciation potential, where opportunities created by dislocation and inefficiencies between and within segments of the real estate capital*

markets are capitalized upon to enhance returns. Investments can include high-yield equity and debt investments and undervalued or impaired properties in need of repositioning, redevelopment, or leasing. Modest leverage is generally applied in value-added portfolios to facilitate the execution of a variety of value creation strategies. (PSERS 2007)

Opportunistic real estate properties are expected to derive a considerable part of their returns from property appreciation and may exhibit substantial volatility in value and returns. Simply put, opportunistic real estate returns are more equity-like. The higher volatility of opportunistic properties relative to the other two styles could be due to a variety of characteristics, such as exposure to development risk, substantial leasing risk, or high leverage, but may also result from a combination of more moderate risk factors, which in total create a more volatile risk profile.

Opportunistic real estate moves away from a core/income approach to a capital appreciation approach. Opportunistic real estate is often accessed through real estate opportunity funds, sometimes called private equity real estate (PERE) funds, which are simply private equity funds that invest in real estate. Given the traditional risk-taking nature of private equity funds, PERE funds often focus on real estate properties with high risk and return profiles, particularly those properties that require extensive development or are turnaround opportunities. Nevertheless, a substantial portion of PERE funds includes value-added and core properties.

The majority of the returns from opportunistic properties come from value appreciation over a three- to five-year period. Rollover risk is high because total return is based on value appreciation. Real estate **rollover** in this context refers to changes in ownership, whereas real estate rollover more generally refers to changes in financing (e.g., converting a construction loan to a permanent mortgage loan) or changes in the nature of a real estate project that facilitate investment liquidity and capability to exit (e.g., completion and full leasing of a project). Since opportunistic properties are held for a shorter term and for their potential to achieve capital appreciation, the risk of failed or delayed rollover is a substantial component of the total risk.

The high rollover risk of opportunistic properties is in contrast to the low rollover risk of core properties. Within the core real estate style, sales of the underlying real estate are infrequent, and properties are held for a long time to harness their income-producing attributes. Due to their high focus on value appreciation, opportunistic real estate managers tend to resemble traders and value enhancers compared to core managers, who are operators of properties. Therefore, opportunistic managers tend to pursue some event (typically a rollover) that will result in the real estate being quickly and dramatically revalued. The capital appreciation of opportunistic real estate can come from development of raw property, redevelopment of property that is in disrepair, or acquisition of property that experiences substantial improvement in prospects through major changes, such as urban renewal.

The investment policy statement of the California Public Employees' Retirement System (CalPERS) identifies opportunistic real estate as follows:

Opportunistic real estate investing is the financing, acquisition, or investment in real estate assets, real estate companies, portfolios of real estate assets, and private and public REITs that do not have access to traditional public equity or debt financing. Opportunistic real estate investing consists *of strategies that seek to exploit market inefficiencies with an emphasis on total return. Opportunistic investments require specialized expertise and the flexibility to respond quickly to market imbalances or changing market conditions. Investments may include nontraditional property types and/or assets that involve development, redevelopment, or leasing risks. Leverage is typically incorporated into this strategy to further enhance total returns. (CalPERS 2006)*

Finally, opportunistic real estate investing is often the way institutional investors expand their property holdings outside their domestic country. Often, institutional investors access cross-border property opportunities through a private equity real estate limited partnership. For example, a public company might decide to sell its investment holdings of apartment housing to private investors, as these real estate properties could represent investments outside the core expertise of the operating companies. This can be the case, for example, of properties that had been held to house the workers of the operating companies. The workers still reside there, but the properties are now in the hands of professional property managers.

16.2 DIFFERENTIATING STYLES WITH EIGHT ATTRIBUTES

The three NCREIF styles can be differentiated using eight major real estate attributes, or characteristics. These attributes were developed by NCREIF to distinguish the three types of real estate asset styles:

1. Property type (purpose of structure, e.g., general office versus specialty retail)
2. Life-cycle phase (e.g., new/developing versus mature/operating)
3. Occupancy (e.g., fully leased versus vacant)
4. Rollover concentration (tendency of assets to trade frequently)
5. Near-term rollover (likelihood that rollover is imminent)
6. Leverage
7. Market recognition (extent that properties are known to institutions)
8. Investment structure/control (extent of control and type of governance)

The styles and their attributes can be used to organize individual properties. Exhibit 16.1 provides descriptions of the three NCREIF styles using the eight attributes of individual real estate properties.

Real estate style analysis can be applied to real estate managers (i.e., portfolios) in addition to individual properties. Exhibit 16.2 provides summary descriptions of the characteristics of real estate portfolios classified into the three NCREIF styles.

16.3 THREE PURPOSES OF REAL ESTATE STYLE ANALYSIS

Real estate styles are essentially locators. In other words, they are categories designed to help identify the space in which each property resides or a real estate manager

EXHIBIT 16.1 The Underlying Eight Attributes of the Three Real Estate Styles

	Core Attributes	Value-Added Attributes	Opportunistic Attributes
Property type	Major property types only: office, apartments, retail, and industrial	Major property types plus specialty retail, hospitality, senior/assisted-care housing, storage, low-income housing	Nontraditional property types, including speculative development for sale or rent and undeveloped land
Life-cycle phase	Fully operating	Operating and leasing	Development and newly constructed
Occupancy	High occupancy	Moderate to well-leased and/or substantially preleased development	Low economic occupancy
Rollover concentration	Tend to be held for a long period of time, forming the central component of the real estate portfolio, which is geared toward generating income rather than sales appreciation	Moderate rollover concentration—a higher percentage of the assets are held for a short- to intermediate-term sale and rollover into new assets	High rollover concentration risk—most of the assets are held for appreciation and resale
Near-term rollover	Low total near-term rollover	Moderate total near-term rollover	High total near-term rollover
Leverage	Low leverage	Moderate leverage	High leverage
Market recognition	Well-recognized institutional properties and locations	Institutional and emerging real estate markets	Secondary and tertiary markets and international real estate
Investment structure/ control	Investment structures often have substantial direct control	Investment structures often have moderate control, but with security or a preferred liquidation position	Investment structures often have minimal control, usually in a limited partnership vehicle and with unsecured positions

operates. There are three main reasons for introducing styles into real estate portfolio analysis:

1. Performance measurement: Investors continually look for tools that can provide them with a better understanding of an investment's or a sector's objectives and success in accomplishing those objectives. This includes identifying peer groups, return objectives, range of risks, return or performance attribution, and

EXHIBIT 16.2 Real Estate Portfolio Style Definitions

Core Portfolio	Value-Added Portfolio	Opportunistic Portfolio
A portfolio that includes a preponderance of core attributes. As a whole, the portfolio will have low lease exposure and low leverage. According to the NCREIF Open-End Diversified Core Equity (ODCE) index for Q4 2014, the long-run average leverage of core funds was 22.2%. A low percentage of noncore assets is acceptable. Such portfolios should achieve relatively high-income returns and exhibit relatively low volatility. The portfolio attributes should reflect the risk and return profile of the NCREIF Property Index (NPI).	A portfolio that generally includes a mix of core real estate with other real estate investments that have a less reliable income stream. The portfolio as a whole is likely to have moderate lease exposure and moderate leverage. According to the NCREIF ODCE index for Q1 2015, the average leverage of value-added funds was 48%. Such portfolios should achieve a substantial portion of the return from the appreciation of real estate property values and should exhibit moderate volatility. A risk and return profile moderately greater than the NPI is expected.	A portfolio predominantly of noncore investments that is expected to derive most of its return from the appreciation of real estate property values and that may exhibit substantial volatility in total return. The increased volatility and appreciation risk may be due to a variety of factors, such as exposure to development risk, substantial leasing risk, high degree of leverage, or a combination of moderate risk factors. A risk and return profile substantially greater than the NPI is expected.

peer performance. Simply put, styles may be useful in identifying appropriate benchmarks.

2. Monitoring style drift: Tracking style drift is another benefit of assessing the style of a portfolio. It is a fact of investing that portfolio managers occasionally drift from their stated risk, return, or other objectives. Classifying different styles of real estate investments allows an investor to assess the association between a portfolio and its underlying investment products as the portfolio changes over time. Identifying the concentration of a portfolio in terms of the styles for each property facilitates a better understanding of the portfolio's risk level at any given point in time.
3. Style diversification: The ability to compare the risk-return profile of a manager relative to the manager's style may allow for a better diversification of the portfolio, since an investor may be able to construct a portfolio that has a more robust risk-return profile if there is a better understanding of each real estate manager's style location. Simply put, style may be useful in understanding and controlling risk.

It should be noted that the preceding real estate styles are primarily applied to private commercial real estate equity, although the concepts can also be applied to publicly traded real estate, such as equity real estate investment trusts (REITs).

16.4 REAL ESTATE STYLE BOXES

The first part of this chapter detailed the use of NCREIF real estate styles to differentiate real estate properties and portfolios by their risks and returns. Chapter 14 discussed numerous other categorizations, including the division of real estate properties by the size of the market in which the property is located. Properties in primary real estate markets such as London are distinguished from properties in secondary (midsize) and tertiary (small) markets. These categorizations can be used to create and use real estate style boxes. **Real estate style boxes** use two categorizations of real estate to generate a box or matrix that can be used to characterize properties or portfolios.

Exhibit 16.3 illustrates style boxes for traditional investments. In the case of the equity style box on the left, the box has equity style on the horizontal axis (e.g., value versus growth) and capitalization size on the vertical axis. In traditional bond analysis, duration is usually on the horizontal axis, with credit quality on the vertical axis.

Style boxes are applied to individual assets, managers, or portfolios. For a style box of an individual stock or bond, the box contains an X in the square most descriptive of the asset. Similarly, managers can be identified with an X in a style box to denote their primary focus. The equity style box in Exhibit 16.3 illustrates the use of an X in a single square to denote the primary characteristic of a hypothetical small-cap growth stock. Portfolios and funds are often identified with percentages in each square denoting how much of the fund's or portfolio's holdings are invested in assets of each location. The fixed-income box on the right of Exhibit 16.3 illustrates the use of percentages in each of the nine squares.

Exhibit 16.4 illustrates real estate style boxes. There is no uniform standard for style boxes in the real estate industry. Clearly, for private commercial equity, the styles of NCREIF are prime candidates for the horizontal axis. Primary, secondary, and tertiary real estate markets are potentially useful for the vertical axis. The left side of Exhibit 16.4 illustrates a potential style box and hypothetical allocations. In this illustration, a real estate style box serves as a method of better understanding the top-down allocations of a real estate portfolio. A real estate style box can also be used to denote the location of a single manager or a single property by placing an X in the relevant square.

Equity Style Box

			Large-cap
			Medium-cap
		X	Small-cap
Value	Blend	Growth	

Fixed-Income Style Box

25%	10%	10%	High credit quality
15%	15%	10%	Medium credit quality
10%	5%	0%	Low credit quality
Short Duration	Intermediate Duration	Long Duration	

EXHIBIT 16.3 Equity and Fixed-Income Style Boxes

Real Estate Allocation Style Box

	Core	Value-Added	Opportunistic
Primary	50%	15%	5%
Secondary	10%	5%	5%
Tertiary	5%	5%	0%

Gross Expected Returns Style Box

	Core	Value-Added	Opportunistic
Primary	$E(R)$ = 7%	$E(R)$ = 7.5%	$E(R)$ = 10%
Secondary	$E(R)$ = 7.5%	$E(R)$ = 8%	$E(R)$ = 11%
Tertiary	$E(R)$ = 8%	$E(R)$ = 9.5%	$E(R)$ = 12%

EXHIBIT 16.4 Real Estate Style Boxes

Note that the box format can be used to organize views of estimated expected returns, volatilities, alphas, and so forth, including metrics discussed in the remainder of this chapter. The right-hand side of Exhibit 16.4 illustrates the concept of using a real estate style box as a tool to organize estimated real estate metrics by inserting arbitrary estimates of expected gross returns (before fees) into a real estate style box.

16.5 CAP RATES AND EXPECTED RETURNS

A somewhat crude but widespread metric in real estate valuation is the capitalization (cap) rate. The cap rate of a real estate investment is the net operating income (NOI) of the investment divided by some measure of the real estate's total value, such as purchase price or appraised value:

$$\text{Cap Rate} = \text{NOI}/\text{Value} \tag{16.1}$$

where NOI is usually viewed on an annualized basis and represents the expected, normalized cash flow available to the owner of the real estate, ignoring financing costs. The value presented in Equation 16.1 is an estimate of the market value of the real estate on an unlevered basis.

APPLICATION 16.5

Assume that a real estate project has a current market value of \$125 million and expected annual cash flows from rent, net of operating expenses, of \$10 million. What is its cap rate?

The answer is found using Equation 16.1: Cap Rate = (\$10 million/\$125 million) = 8%.

The exact specifications of the NOI for the numerator of the cap rate (recent, current, forecasted) and the value (beginning of period versus end of period, transaction price versus appraised price) for the denominator of the cap rate vary between users and purposes. Note that NOI does not reflect financing costs, and therefore the NOI-based approach for estimating value depicted in Equation 16.2 is intended for analysis of unleveraged property values.

Cap rates are often viewed as direct estimates of expected returns. Thus, a property with a cap rate of 9% is expected to generate a return of 9% to the investor on an unleveraged basis. The view of a cap rate as an estimated expected return is at best a crude approximation in that it typically ignores anticipated capital gains or losses as well as anticipated growth or decline in income. Nevertheless, cap rates are a good starting point for an analysis of expected returns.

Cap rates are also viewed as required rates of return and are used to perform risk adjustment in valuation of properties. Thus, an investor may search for a core property that offers a cap rate of at least 7% while demanding a cap rate such as 9% on a value-added property that is perceived as having higher risk.

Cap rates are often used to establish values for particular properties. Equation 16.2 rearranges Equation 16.1 to express property value as depending on NOI and cap rate:

$$\text{Value} = \text{NOI}/\text{Cap Rate} \tag{16.2}$$

A prospective buyer or analyst divides the estimated NOI of a property by an industry standard cap rate to obtain an estimate of the property's unleveraged value. For example, an investor considering the purchase of a property that offers \$350,000 per year in net operating income may decide that if cap rates on comparable properties are 7%, then the property is worth \$5,000,000.

Other metrics that are commonly used to assess real estate investment opportunities mirror the approaches discussed in the private equity section, including the internal rate of return (IRR) approach and the multiple of the cash in (the cash projected to be received from an investment on a project) to the cash out (the investment in the project). Numerous ratios of operating performance, interest coverage, and leverage are also used, most of which are analogous to financial ratios used throughout corporate finance. Cap rates represent a tool that is widely used in real estate but less so in other areas of finance.

NCREIF and other organizations estimate cap rates for various commercial property types in the United States. The overall weighted average of cap rates for core commercial property in the United States has varied between 5% and 10% since the mid-1980s. The higher end of the range was reached in the mid-1990s, whereas the lower end of the range was a result of the onset of the financial crisis that began in 2007 and ended around 2009. Estimates of recent cap rates on U.S. commercial property (in early 2015) ranged slightly below long-term averages (at between 5% and 6%) for core real estate properties in primary real estate markets.

16.6 DEVELOPING RISK AND RETURN EXPECTATIONS WITH STYLES

There are two major levels at which estimation of risk and return involving real estate should be performed by an asset allocator: (1) at the categories or styles level, and

(2) at the individual fund, manager, or property level. The investor will then compare the risk and return of prospective funds, managers, or properties with the risk and return required or anticipated for the associated categories.

In traditional investments such as equities, an analyst's expectations of risk and return may be predominantly driven using measures of risk derived from observation of historical market returns and prices. Thus, measurement of the risk of a particular stock (or a category of stocks, such as midsize domestic equities) may be derived from observation of the volatility and correlation of its past returns. However, much of real estate equity involves private real estate, with little or no direct and consistent observations of market prices.

The unique nature of each real estate property and the general lack of liquidity make the tracking of real estate returns and the formation of risk and return estimates more difficult than they are for publicly traded asset classes. However, when using styles, it is important to have some range of risk and return expectations in order to value properties and to determine how a real estate property or investment manager should be monitored and evaluated. In the context of real estate styles, the challenges of estimating risk and return for the value-added and opportunistic segments of the market are daunting, while the process for core properties is less difficult.

The key to developing risk and return estimates for real estate styles is to begin by estimating the expected risks and returns of the core real estate properties, and then to express the estimates for the value-added and opportunistic segments relative to the core style. Thus, an analyst may use cap rates or historical returns to estimate that the expected return of core real estate properties is, for example, 8%, and then estimate the expected returns of value-added and opportunistic real estate by approximating a premium by which those segments exceed the expected returns of the core segment.

16.6.1 Core Real Estate Expected Return

Development of expected risk and return estimates for core real estate properties can be based primarily on past returns and current metrics, such as cap rates. The NCREIF Property Index (NPI) provides an appraisal-based historical database of core property returns. The 25 years (100 quarters) ending on December 30, 2014, had a mean quarterly return of 1.90%. The quarterly inflation rate (measured using the Consumer Price Index) was 0.61% during this period.

As detailed in Chapter 15, the NPI is generally believed to be a smoothed index due to the use of appraised values. Despite the problems with smoothed data discussed in Chapter 15, data smoothing would typically have only a small effect on long-term historical mean returns. Any effect of smoothing on mean returns emanates only from the extent to which the starting and ending valuations are smoothed. Over a 25-year observation period, even major valuation errors in starting and ending values have minor effects on mean returns. For example, if the pricing error due to smoothing for the initial valuation is 10% higher than the final valuation, a 25-year annual mean return would be biased downward by only 10%/25 = 0.4% (ignoring compounding).

Note further that consistent pricing errors may have little or no effect on mean returns. If the index consistently overstates (or understates) the values of the underlying properties, then the returns of the index will match the returns of the true market values. For example, if the index values are consistently 10% lower than the true values, the mean of the returns of the index will be the same as the mean returns of the true prices (ignoring income).

As the observation interval is lengthened, the effects of valuation errors, including smoothing, on the estimated mean return of an index are decreased, and the estimated mean return approaches the true mean return. Thus, the 25-year historical compounded annualized mean return for unleveraged core real estate of 7.8% is probably a reasonable estimate of the average returns experienced in the United States by core real estate investors over that period. Of course, the question remains as to whether the mean of past returns is indicative of the expected future returns.

16.6.2 Core Real Estate Risk

The 25 years of quarterly returns of the NPI ending on December 30, 2014, had a quarterly standard deviation of 2.4%. Without adjustment for smoothing, the implied annualized volatility is 4.8% (found by multiplying the periodic standard deviation by the square root of the number of the periods in one year—in this case, the square root of 4).

Smoothing of returns interferes far more in risk estimation than in long-term return estimation. The smoothing of returns can severely dampen the estimated volatilities and correlations of returns. Thus, the estimated standard deviation of the quarterly returns of the smoothed series (2.4%) needs to be adjusted for the effects of smoothing on implied annual volatility of true real estate values. The degree of volatility dampening depends on the degree of return smoothing. Equation 16.3 expresses a rough approximation of the volatility of a return series that has been smoothed with first-order autocorrelation, as described in Chapter 15:

$$\text{True Volatility} = \text{Smoothed Volatility} \times \sqrt{\frac{1+\rho}{1-\rho}} \tag{16.3}$$

where ρ is the first-order autocorrelation of the smoothed series. It is assumed that the underlying true series has no autocorrelation.

APPLICATION 16.6.2a

Use an arbitrary estimate of general real estate return autocorrelation of 0.60 to estimate the true volatility of a return series with a smoothed volatility of 2.4%. First, solve for the square root term in Equation 16.3: $\sqrt{(1+\rho)/(1-\rho)} = 2$. Applying Equation 16.3 to unsmooth the NPI's return volatility increases the quarterly return volatility from 2.4% to 4.8% (found as 2.4% × 2), which annualizes to 9.6% volatility. However, the NPI appears to have an autocorrelation coefficient higher than 0.6, so the unsmoothed volatility could be much higher. Note also that Equation 16.3 can be factored as: Smoothed Volatility = True Volatility ÷ $\sqrt{(1+\rho)/(1-\rho)}$. Thus, a smoothed returns series with first-order autocorrelation of 0.60 could roughly be expected to exhibit only 50% of the volatility of the true returns series.

Estimation of systematic risk using smoothed historical data is even more problematic than the estimation of volatilities. Equation 16.4 denotes the relationship between systematic risk measures based on true returns and smoothed returns:

$$\beta_{\text{true series}} = \beta_{\text{smoothed series}}/(1-\rho) \tag{16.4}$$

where $\beta_{\text{true series}}$ is a systematic risk measure based on the true return series underlying the smoothed series and a particular market index, $\beta_{\text{smoothed series}}$, is a systematic risk measure based on a smoothed return series and a particular market index, and ρ is the first-order autocorrelation of the smoothed series.

APPLICATION 16.6.2b

Assume that a smoothed return series has β equal to 0.20 and that the true but unobservable return series has β equal to 0.50. What would the estimated autocorrelation of the smoothed series be?

Inserting the two betas and solving for the correlation coefficient generates an estimated autocorrelation coefficient of 0.60.

The implication in the case of both systematic risk estimation and volatility estimation is that a smoothed return series exhibits dramatically lower risk than is contained in the true return series. Chapter 15 demonstrated an unsmoothing procedure that can be used to estimate the underlying true return series and its risks. However, the accuracy of the unsmoothing procedure depends on the accuracy of the model specifying the smoothing (e.g., first-order autocorrelation) as well as the accuracy of the estimation of the parameters of the model (e.g., the autocorrelation coefficient). Substantial errors in specifying the unsmoothing model and estimating its parameters can generate large errors in the estimation of the risks of the true underlying assets.

16.6.3 Return Estimates for Noncore Assets and the Risk-Premium Approach

While data on core real estate returns are reasonably available and reliable, data on value-added and opportunistic risks and returns are more problematic.

Accordingly, the expected risks and returns of value-added and opportunistic real estate investments are typically estimated and expressed relative to the risks and returns of core real estate investments. Risk-premium methodologies can serve as effective tools both in estimating the current expected returns of core properties and in developing expected returns for value-added and opportunistic properties.

As previously discussed, expected returns from real estate are often assumed to be approximately equal to cap rates. Cap rates are often expressed as a cap rate spread. The **cap rate spread** is the excess of the cap rate over the yield of a default-free 10-year bond (such as the 10-year Treasury rate in the United States). A typical cap rate spread for core real estate over the 10-year Treasury rate is 2% to 3%, or 200 to 300 basis points.

In the finance literature, expected investment returns are often modeled using a risk premium approach, which expresses a risky asset's return as the sum of a riskless return and a premium for bearing the risk of that asset. The risk premium approach is illustrated in Equation 16.5:

$$E(R_i) = R_f + \text{Risk Premium}_i \tag{16.5}$$

where $E(R_i)$ is the expected return on asset i, R_f is the riskless interest rate, and risk premium$_i$ is the risk premium, or spread, of asset i relative to the riskless rate. Note that risk premium$_i$ is a premium peculiar to asset i; it is not a risk premium such as the premium on the market portfolio that is used to determine cross-sectional expected returns of other assets based on their correlation with the market.

The key contribution of the risk premium approach in Equation 16.5 is its potential to incorporate the current riskless interest rate in forming expectations of future returns for the same asset for different market conditions. Since riskless interest rates vary primarily due to inflation expectations, Equation 16.5 may be viewed as a method of adjusting return expectations for different anticipated inflation levels as well as different interest rate levels. For example, in the late 1970s, U.S. inflation rates reached double-digit levels. It would be unreasonable to believe that investors expected single-digit nominal returns on competitively priced assets when there was double-digit inflation. Nominal after-tax interest rates must exceed expected inflation rates for real rates to be positive.

The maturity of the default-free interest rate used in the risk premium approach varies and is generally determined by the investment horizon of the application. Outside of real estate, analysts often use a very short-term rate, such as an overnight rate or a three-month rate, to capture the rate earned in the absence of default risk and interest rate risk. In U.S. real estate markets, the 10-year Treasury rate is typically used as the default-free interest rate against which to evaluate cap rates and risk premiums. Presumably, the 10-year bond is selected to more closely approximate the longevity of real estate holdings. The decision of which riskless interest-rate maturity to use is a matter of professional judgment.

The U.S. interest rates on 10-year-maturity Treasuries over the same 25-year period used previously to estimate the NPI returns averaged approximately 4.90%, and average interest rates on one-year-maturity Treasuries were approximately 3.3%.

APPLICATION 16.6.3

Solve for the average risk premiums of the NPI given a 7.8% NPI average annualized return and the one-year and 10-year Treasury average returns of 3.3% and 4.9%.

Using Equation 16.5 generates 25-year average risk premiums of 4.5% and 2.9%, respectively.

Note, however, that in 2011 and 2012 cap rates on core real estate, perhaps in the area of 7%, exceeded the 10-year Treasury rate (which ranged around 2% to 3%) by 4% to 5%, well out of line with average historical experience for core real estate (2% to 3%). Perhaps these observed cap rate spreads over 10-year Treasuries in 2011 and 2012 can be explained by the extraordinarily low Treasury yields associated with the U.S. Federal Reserve's attempts to help the U.S. economy emerge from the global financial crisis. Alternatively, in view of the extreme volatility in real estate prices since the onset of the global financial crisis in 2007, it may be reasonable to argue that the recent estimated risk premiums exceed historical risk premiums to compensate prospective real estate investors for the unusual levels of risk. A final possibility is that the market believed that current estimates of net operating income, capital appreciation, or both were too high.

Development of an estimate of the expected returns of value-added and opportunistic real estate investing may take three primary approaches: (1) the use of observed cap rates for those styles of real estate investment, (2) the use of a risk premium approach (based on the estimation of their risks relative to the risks of core real estate investments), or (3) the use of absolute hurdle rates.

The direct use of observed cap rates in market transactions to estimate expected returns for value-added and especially opportunistic properties may be rather inaccurate because the estimations of net operating income for these styles are much less reliable than are the estimations of NOI for the core style. The use of absolute hurdle rates may be inappropriate due to their failure to consider different interest rates and expected inflation levels. Accordingly, it may be argued that the best estimate of expected returns on noncore properties can be found by adding a risk premium to the expected returns estimated for core properties.

16.6.4 Examples of Return Estimates for Noncore Style Assets

At particular points in time, investors may develop absolute target rates of return for properties with value-added and opportunistic styles rather than relying on observed cap rates or cap rate spreads. This section reviews two examples of the setting of returns by institutions.

PSERS (2014) defines its opportunistic return target in absolute terms. It sets a hurdle rate for opportunistic real estate investing at 16% or greater, depending on the level of risk taken. CalPERS (2014) also defines its opportunistic return target in absolute terms. It cites an expected hurdle rate of 13% or greater. The CalPERS investment policy goes further to note that investment staff may adjust this 13% hurdle rate depending on the characteristics of the individual opportunistic real estate fund or changes in the marketplace, including changes to the inflation rate, capital market risk levels, or levels of available investment opportunities. The CalSTRS investment policy for opportunistic real estate (2014) provides additional guidance:

> *Investments within the opportunistic portfolio seek to capitalize on tactical opportunities, mispricing, or distress in the real estate and capital markets and are willing to assume additional risk. Investments include direct real estate assets in office, retail, industrial, residential, or specialized property types, as well as forms of investment such as land plays, operating companies, distressed debt/properties, and other specialized investments (e.g., brownfields). Opportunistic investments utilize development, extensive redevelopment, nontraditional investment vehicles (e.g., non-performing loans), entity level investments, or recapitalization of assets or companies. Investment strategies may also include international/emerging markets. (CalSTRS 2014)*

The last part of the CalSTRS policy mentions investment strategies in international/emerging markets because many institutional investors look at real estate investing outside of their home country as being opportunistic in nature, since it involves a market in which real estate is valued differently, in which development issues such as planning and zoning can be much more difficult, and in which property management must recognize the peculiar nature of the foreign market. So, for example, when an institution invests overseas, it needs to understand the laws regarding foreign taxation, as the tax-exempt status in the home country may not be recognized abroad, and it must deal with the conversion of the total return back into local currency. These are additional risk factors for which additional expected returns must compensate.

One of the difficulties of assessing the return expectations for opportunistic real estate is that many of these investments take place through private limited partnerships (i.e., private equity real estate). Because these limited partnerships are private, returns tend to be based on appraised values rather than market values.

16.7 CHARACTERISTICS OF REAL ESTATE DERIVATIVES

This section describes real estate derivatives and their use by institutional investors to hedge against or to speculate on real estate assets. Derivative products allow investors to cost-effectively transfer risk exposure related to either the equity or the debt sides of real estate investments without having to actually buy or sell properties. This is accomplished by linking the payoff of the derivative to the performance of a real estate return index, thus allowing investors to obtain exposures without engaging in real estate property transactions or real estate financing.

Tradable real estate derivatives based on indices were developed at the start of the 21st century, beginning in the United Kingdom (derivatives based on the Investment Property Databank index) and later in the United States (derivatives based on the NPI). Furthermore, the CME Group—the combined entity created by the

2007 merger of the Chicago Mercantile Exchange (CME) and the Chicago Board of Trade (CBOT)—began offering futures contracts based on the S&P/Case-Shiller Home Price Indices for major U.S. cities, as well as a composite index of all the cities. These derivative products suffered a severe blow as a result of the real estate and financial crisis of 2008–9.

16.7.1 Three Benefits of Housing Price Derivatives

The emergence of derivatives on housing prices offers three major potential benefits: (1) price revelation, (2) better risk management, and (3) the ability to short-sell residential real estate. Note that, unlike major securities markets for stocks and bonds, the cash market for residential real estate provides virtually no opportunity to establish short positions.

16.7.2 Two Critical Factors for Effective Risk Management with Derivatives

The value of real estate derivatives to offer effective risk management tools to home builders, developers, and investors depends on two critical factors: (1) The index on which the derivatives are based must accurately represent the changes in the value of the underlying cash market, and (2) the derivatives must offer substantial liquidity. S&P/Case-Shiller Home Price Indices and their futures contracts are examples of the types of innovation that emerged to bring advanced risk management tools to all areas of equity real estate investing.

16.7.3 Baum and Hartzell's Seven Advantages of Property Derivatives

According to Baum and Hartzell (2012), property derivatives have the following seven advantages:

1. Low transaction costs compared to direct investments in real estate, which can incur round-trip costs up to 10%
2. Instant access to real estate returns
3. Relatively small management costs compared to those of direct real estate investments
4. Diversified exposure to real estate returns
5. Hedging and other strategies by allowing participants to take short positions on the real estate market
6. Not currently subject to withholding taxes
7. Leverage without needing debt

16.7.4 Baum and Hartzell's Six Disadvantages of Property Derivatives

According to Baum and Hartzell (2012), property derivatives have the following six disadvantages:

1. The real estate derivatives market is still relatively new and in many cases illiquid.
2. Cash flows from real estate derivatives are often subject to significant counterparty risk. This risk, which varies depending on the type of contract and the credit risk of the counterparty, can be hedged, but it incurs an additional cost.
3. Acceptable real estate indices, which are the base of real estate derivatives, are still not available in many markets, particularly in emerging markets.
4. Highly leveraged speculative strategies could lead to large losses and thus wipe out margin balances.
5. Basis risk exists between the performance of a real estate portfolio and the underlying index.
6. Real estate derivatives are marked-to-market on a monthly basis. This low frequency creates significant accounting volatility.

16.8 TYPES OF REAL ESTATE DERIVATIVES AND INDICES

This section discusses types of real estate derivatives and specialized real estate indices.

16.8.1 Property Total Return Swaps

In a **property total return swap**, the buyer of property exposure agrees to pay a fixed price each year over the life of the contract, and in return, the swap seller agrees to pay the annual total return of the applicable property index. Only the net cash flows are exchanged between the parties, and no cash is exchanged at the beginning of the life of the contract. This implies that these contracts have a relatively low counterparty risk. Property total return swaps are traded over the counter. Margins are determined by the two parties' credit risk and typically represent between 5% and 20% of the notional value of the contract. The following types of property swaps are available: single (specific) property swaps, property sector swaps, multi-asset level swaps, and international index swaps.

Total return swaps based on commercial property indices were the most popular property derivative in the United Kingdom until the time of the global financial crisis of 2008–9. The basic UK swap was an over-the-counter contract in which parties swapped an annual Investment Property Databank (IPD) property index total return for that country for an established interest rate. A similar product was launched in the United States based on the NPI, and it was also a casualty of the global financial crisis of 2008–9. In spite of these negative developments, real estate total return swaps are regarded as having strong potential for the future, and therefore it is sensible to study them here.

According to Baum and Hartzell (2012), market participants consider the fixed rate of a swap to represent the implied expected return on that specific segment of the real estate market. Property total return swaps can be used in different ways in portfolio management. For example, they allow investors predicting a downturn in the real estate market to hedge their physical real estate holdings by selling market exposure for a certain period of time while preserving the ownership of real estate assets. Also, swaps can offer instant exposure to real estate assets (assuming that

EXHIBIT 16.5 Example of a Property Total Return Swap

Year	RE Index	Floating Rate	Difference	Present Value
1	9.2%	6.00%	$3,200,000	$3,018,868
2	6.4%	5.15%	$1,250,000	$1,112,496
3	7.2%	4.45%	$2,750,000	$2,308,953
			Total	**$6,440,317**

a suitable index exists) compared to the lead time necessary to invest directly in real estate.

Example: Assume that a three-year floating-rate swap with a notional contract value of $100 million is to be negotiated. Let us suppose that a buyer of a real estate index forecasts that real estate returns (as reflected by "RE Index," a hypothetical real estate index ideally including both income and capital appreciation) will be 9.2%, 6.4%, and 7.2% during the next three years. The floating interest rate (e.g., LIBOR) will be 6.00%, 5.15%, and 4.45% during the next three years. Assume that no margin above the floating rate was paid for this swap. Finally, suppose that the required return for investing in property (which reflects the risk-free rate plus a risk premium) is 6% per year for the next three years.

What is the present value of the expected payments to be paid/received by the buyer of the RE Index in this swap during the next three years? Exhibit 16.5 shows the calculations of this total return swap, in which the investor will receive yearly the difference between the return on the RE Index and the floating rate multiplied by the notional contract value of $100 million. For the first year this amounts to $3.2 million: (9.20% – 6.00%) × $100 million = $3.2 million. The present value of this expected payment is $3,018,868 (i.e., $3,200,000/1.06). The total present value of the expected payments for the three years is $6,440,317. Note that the market value of a swap at the time that the swap is struck is zero. Obviously, the buyer in this swap is more optimistic about the performance of real estate assets (as measured by the RE Index) than the rest of the market.

APPLICATION 16.8.1

Consider a three-year floating-rate swap with a notional contract value of $50 million. Suppose that a swap buyer forecasts that real estate returns will be 9% per year, while LIBOR will be 6% during the next three years. Using a discount rate of 5%, what is the present value of the expected payments to be paid/received by the buyer of this swap during the next three years?

The expected cash flows will be (9% – 6%) × $50 million, which equals $1.5 million per year. A three-year $1.5 million annuity discounted at 5% is worth $4.08 million. Note that an application of this topic could provide the present value and most other data and require computation of the missing value.

Observe that if the difference between the return on the RE Index and the floating rate were to be negative for a certain year, the investor would have to pay that difference multiplied by the notional contract value of $100 million. A swap contract could also be settled at another frequency—for example, quarterly—provided that the real estate index underlying the swap is updated and reported quarterly.

An index buyer on a swap agreement can construct a synthetic investment in a real estate total return index. For example, consider the case of an endowment fund interested in making a diversified investment of $20 million in the U.S. institutional real estate market following a pre-established portfolio allocation policy to the real estate asset class. The endowment will invest $20 million in one-year U.S. government securities and use the interest proceeds from this investment to cover its payment obligations under the swap, in which the endowment is an index buyer on a broadly diversified index reflecting the U.S. institutional real estate market. The index is calculated and reported quarterly. Assuming that this is a fixed-rate swap and that the fixed payments to be received from the riskless bond investment equal the fixed payments to be made to honor the swap agreement, the net cash flow from the combined position will generate exactly what the total return would be each quarter on a $20 million investment on the real estate index. Thus, the endowment is exactly assuming the risk and return of the U.S. real estate index without the need to make direct investments in real estate. This operation is known as a covered long position in the swap.

In the previous example, there could be, for example, a private equity real estate fund as the index seller under the swap contract. Suppose that the fund has an exposure to the U.S. institutional real estate market in excess of its portfolio allocation target of $20 million. By selling the real estate index under the swap, the private equity fund will have to make quarterly payments that will be expected to closely match the cash flow that the fund receives each quarter from its $20 million investment in real estate assets. Therefore, the swap creates a covered short position for the fund.

16.8.2 Forwards, Futures, and Options Contracts

Exchange-traded contracts such as real estate futures and options have the advantage that they guarantee that all parties have deposited the mandatory collateral; therefore, counterparty risk is mitigated considerably.

Since February 2009, Eurex—the international derivatives exchange—has listed property index futures. The futures contract is based on the total returns of the IPD UK Annual All Property Index. The exchange lists five consecutive annual contracts with pricing based on a par value of 100 plus expected percentage total annual return in the related calendar year. For example, the contract for the calendar year 2009, which expired in March 2010 (expiry is the last working day in the following March to ensure it is after publication of the IPD data), settled at 103.50, representing a +3.5% annual total return (as published by IPD). In 2014, a total of 5,474 contracts traded according to the Eurex website, representing EUR 370 million in capital value. As of June 2015, there were plans to launch new futures contracts based on IPD's property indexes (e.g., UK subsector indexes) and other European property indexes.

In the United States, property derivative trading is primarily made through forward and futures contracts. Forward agreements are made generally on the RPX and

NCREIF indices. Futures and options trading is done by the CME Group via Globex utilizing the S&P/Case-Shiller Home Price Indices (CSI). CSI futures are cash-settled to a weighted composite index of U.S. housing prices (Boston, Chicago, Denver, Las Vegas, Los Angeles, Miami, New York, San Diego, San Francisco, and Washington, DC). Each contract is valued at $250 times the CSI for that city. For example, if the value of the index for San Diego was reported at 300, the contract value would be $75,000 (i.e., $250 × 300 = $75,000). The CSI futures contracts trade for the next 18 months and are listed on a quarterly cycle. All contracts are settled in cash on the day the values of the indices are released. The tick value for a CSI futures and options contract is 0.20 of an index point. Thus, the value of a tick is $250 times 0.20 = $50.00 per contract. Therefore, if the current price of the contract is 263.30, a positive move of one tick would take the price of the contract to 263.50, and the contract would rise in value by $50: (263.50 – 263.30) × $250 = $50. The long position would be credited $50, and the short position would be debited $50. All positions are marked-to-market twice each day.

There are also CSI options on futures. The contract size is one futures contract, and the minimum price fluctuation is 0.10 index points, which is equivalent to $25 per contract. These options contracts are exercised into the associated futures contract.

16.8.3 Real Estate Index Notes

Structured real estate index notes are bonds whose coupons are associated with the performance of a specific real estate index. The note issuer is paid a capital sum at the beginning of the contract, while the owner of the note receives an income arising from the total return on a property index, with a return of principal at the maturity of the contract.

Real estate index notes are similar to swaps. However, and unlike swaps, real estate notes are funded products, and therefore cash is exchanged at the beginning of the contract, causing these products to have greater counterparty risk than total return swaps. In the United Kingdom, property index notes (PINs) are an example of a real estate index note.

Real estate index notes can be regarded as an investment in a plain-vanilla bond and a total return swap. For example, consider the case of a pension fund desiring to increase its exposure to real estate. The fund purchases a real estate index note for $10 million with the expectation of receiving the performance of the respective real estate index throughout the life of the contract, and the capital is returned to the pension fund at the end of the life of the contract. Notice that this position is equivalent to purchasing a bond at par value with $10 million in face value and taking a long position in a total return swap.

16.8.4 Traded Derivatives of Mortgage-Backed Securities

Derivatives based on the CMBX—indices of commercial mortgage-backed securities (CMBS)—consist of a family of indices encompassing 25 CMBS with various credit ratings. These indices let participants, by investing in credit default swaps, to hedge against or to speculate on risks that follow the credit strength of the commercial mortgage market.

The ABX index (subprime residential mortgage-backed securities [RMBS]) is a family of credit default swaps. It is based on 20 bonds that consist of subprime mortgages. ABX contracts are commonly used by market participants to hedge against or to speculate on the risk that the underlying mortgage-backed securities will not be repaid as originally thought. ABX swaps offer regular insurance-like premiums if the underlying mortgage-backed securities are not repaid as expected.

16.8.5 Stock Market–Based Property Return Indices

Stock market–based property return indices (SMPRIs) track property-level returns based on price changes of stocks that focus on commercial property investment (e.g., commercial property REITs in the United States). To achieve this, the share prices of REITs are delevered, and the resulting movements of these "pure play" stocks are then used to track how the stock market's valuation of the properties held by the REITs changes.

The first of these indices is the FTSE NAREIT PureProperty Index, which was launched in the United States in 2012. This index is updated daily (compared to indices such as the NCREIF Transaction Based Index [TBI], which is reported only quarterly), and its construction is objective and replicable. Geltner et al. (2014) show that the trends exhibited by the FTSE NAREIT PureProperty Index between 2000 and 2012 were very similar to those corresponding to private real estate market trends (TBI and NPI), with the SMPRI sometimes leading the private market indices. Together with the daily pricing, these high correlations render SMPRIs to be, at least in theory, suitable and very promising potential candidates for serving as the basis for derivative products.

16.9 CONCLUSION

This chapter has discussed a common style classification used by institutional investors to gain better understanding of their real estate investments. As we move from core to value-added and then to opportunistic real estate properties, we observe investment products with higher risk and higher expected returns. Core properties behave like bonds because of their long-term and steady cash flows. Opportunistic properties behave like growth stocks because their total returns are driven mostly by price appreciation.

The chapter also discussed real estate derivatives and their use in portfolio allocation and risk management. Real estate derivatives may represent a quick and convenient way for institutional investors to gain exposures to real estate in the short run. In addition, such derivatives allow investors to take short positions in real estate, thus reducing their overall exposure to this asset class.

REFERENCES

Baum, A. E., and D. Hartzell. 2012. *Global Property Investment: Strategies, Structures, Decisions*. West Sussex, UK: Wiley-Blackwell.

CalPERS. 2014. "Comprehensive Annual Financial Report: Fiscal Year Ended June 30, 2014." www.calpers.ca.gov/docs/forms-publications/cafr-2014.pdf.

CalPERS. 2006. "Investment Policy Statement."

CalSTRS. 2014. "Real Estate Investment Policy." April. www.calstrs.com/sites/main/files/file-attachments/k_-_real_estate_investment_policy.pdf.

Geltner, D., N. Miller, J. Clayton, and P. Eichholtz. 2014. *Commercial Real Estate Analysis and Investments*. Mason, OH: OnCourse Learning.

PSERS. 2007. "Comprehensive Annual Financial Report: Fiscal Year Ended June 30, 2007."

PSERS. 2014. "Comprehensive Annual Financial Report: Fiscal Year Ended June 30, 2014." www.psers.state.pa.us/content/publications/financial/cafr/cafr14/Combined%202014%20CAFR.pdf.

CHAPTER 17

Listed versus Unlisted Real Estate Investments

A primary challenge of real estate investment management is to assess the relative reliability of the conflicting indications of risk generated by observations of publicly traded (or listed) real estate investments, such as real estate investment trusts (REITs), and privately held (or unlisted) real estate. Privately held real estate is typically highly illiquid, taking months or even years to trade at competitive prices. This illiquidity is driven by the extent to which each property is unique. The high illiquidity of most real estate can raise serious challenges for portfolio management, including cash management, risk measurement, and risk management. Listed real estate is highly liquid, with readily observable market prices. But indices of listed real estate prices indicate substantially higher volatility and substantially lower diversification benefits than do indices of private real estate.

This chapter examines the main unlisted real estate funds available (open-end funds and closed-end funds), as well as the most important listed real estate funds (REITs and exchange-traded funds [ETFs] based on real estate indices). The chapter also analyzes the extent to which analysis of publicly traded real estate may be used to provide information on the risks and returns of private real estate, and offers an empirical analysis of the differences between the returns of listed (market-based) real estate and those of private (appraisal-based) real estate.

17.1 UNLISTED REAL ESTATE FUNDS

This section presents unlisted real estate funds. These include open-end and closed-end real estate funds, as well as funds of funds. The section also discusses the most important advantages and disadvantages of privately held real estate funds.

17.1.1 Open-End Funds

Open-end real estate funds allow investments and redemptions (usually after an initial lockup period) at any time. Open-end funds have an indefinite life. The redemption price is calculated based on the quarterly appraisal value of the portfolio of properties. This can create serious problems; for example, when real estate prices are increasing, a rush of buyers will invest in the fund at the last quarterly price, which does not yet entirely reflect the upturn in the real estate market. The resulting new cash will dilute the real estate return provided by the manager and hurt her performance in the short run, while possibly increasing the fund's net asset value (NAV) in the long run. In a declining real estate market, a rush of sellers trying to benefit from the relatively high appraisal value of the previous quarter will hurt the other investors in the fund as well as the manager's performance.

U.S. open-end real estate funds are typically required to fund redemptions out of the income received from the underlying portfolio of properties. If there is not enough earned income to fund redemption requests, investors wanting to redeem their shares must wait until other investors in the fund (who were first in asking to redeem their shares) are paid. These restrictions are not usually present in European open-end real estate funds.

In the case of the UK, **property unit trusts (PUTs)** are the main open-end investment product used by pension funds and insurance funds to obtain a diversified exposure to the UK real estate market. The prices of these unlisted investment vehicles are calculated using the appraisal valuation method.

Unauthorized PUTs are unregulated unit trusts that may be offered only to institutional investors. All issued units are purchased by these investors, who are completely exempt from capital gains tax or corporation tax (e.g., pension funds and charities). **Authorized PUTs (APUTs)** are intended mainly for retail investors and offer exemption from capital gains tax on disposals of investments in the fund. Furthermore, there is no extra tax liability for exempt or corporate investors on distributions from the fund, but credit for any tax paid into the fund is not available. According to Baum (2010), APUTs are not particularly attractive to exempt funds because the structure has an absolute tax cost that they can circumvent by investing in an unauthorized PUT. Finally, offshore PUTs are tax-effective for many UK and international institutional investors who are either tax-exempt or non-exempt.

Property authorized investment funds (PAIFs) are investment vehicles authorized by the UK's Financial Conduct Authority that can invest in real estate directly or indirectly (mainly though shares in UK REITs). The Financial Conduct Authority requires PAIFs to value their positions on a daily basis. APUTs can be converted to PAIFs if certain conditions are met.

17.1.2 Closed-End Funds

Closed-end real estate funds issue a general number of shares to investors before any real estate investments are made. Afterward, investors cannot liquidate their positions, instead redeeming their shares or units or trading them in the secondary market, which tends to be quite illiquid (and hence our decision to categorize this investment vehicle as an unlisted real estate product). In the United States, shares or units are tradable only on a matched-bargain basis. Closed-end real estate funds have a specific investment period (e.g., three years) and fund-termination period (e.g., four to six years before the termination period concludes), at which time the fund must distribute all cash flows to investors.

In the United States and internationally, the majority of closed-end real estate funds follow the format of the limited partnership. One of the main advantages of limited partnerships is that they are tax-neutral investment vehicles. The closed-end real estate funds just described should not be confused with closed-end real estate *mutual* funds (CEMFs), which are exchange-traded mutual funds that have a fixed number of shares outstanding and that represent only a small portion of the listed real estate industry. CEMFs issue a fixed number of shares to the general public in an initial public offering; in contrast to open-end mutual funds, shares in closed-end mutual funds cannot be obtained from or redeemed by the investment company. Instead, shares in closed-end real estate mutual funds are traded on stock exchanges; therefore, they should be classified as a listed real estate investment. CEMFs invest in REITs and real estate operating companies and offer the possibility of gaining a diversified exposure to the real estate asset class with just a small amount of capital. Whereas shares on closed-end real estate funds are fairly illiquid (and, as mentioned previously, are tradable only on a matched-bargain basis in the United States), shares on CEMFs are generally liquid.

Because shareholders cannot create or redeem CEMF shares directly with the fund, arbitrageurs cannot continuously keep closed-end mutual fund prices at or near their NAVs. Like other closed-end mutual funds, real estate CEMFs often trade at premiums or discounts to their NAVs, especially when the NAVs are not based on market values. Real estate closed-end mutual funds usually liquidate their real estate portfolios and return capital to shareholders after an investment term (typically 15 years), the length of which is stated at the fund's inception. Closed-end real estate mutual funds invest in properties, REITs, or both, and experienced substantial growth up until the time of the real estate crisis of 2007–8.

Finally, Exhibit 17.1 summarizes the main characteristics of open-end and closed-end real estate funds as they pertain to the following: number of funds as a percentage of total unlisted finds, types of real estate assets that they hold, life of the fund, liquidity, redemption price, fund structures, tax status, and fees.

17.1.3 Funds of Funds

Real estate funds of funds invest in other real estate funds rather than investing directly in real estate assets. Real estate funds of funds provide higher diversification than investing in only a few real estate funds, and are run by managers who are specialists in the real estate arena. The downside to investors in real estate funds of funds is that they are charged two fees: the first by the fund of funds manager, and the second by the managers of the underlying real estate funds. Real estate funds of funds in the UK (and the United States) are open-ended, whereas in most of the rest of Europe they are closed-ended.

17.1.4 Four Advantages of Unlisted Real Estate Funds

Four advantages of unlisted real estate funds are:

1. They help diversify real estate specific risk. Large sums of money are required to build a diversified portfolio of real estate physical assets. Baum and Struempell (2006) report that more than £1 billion is required to form a diversified portfolio of offices in London. Assuming a structure of 50% debt and 50% equity financing, 20 investors committing £25 million each would be required to raise the needed capital to form an unlisted real estate fund. However, £25 million can afford the purchase of only one or two offices of average lot size in London.

EXHIBIT 17.1 Main Characteristics of Unlisted Real Estate Funds

	Open-End Real Estate Funds	Closed-End Real Estate Funds
Percentage of total unlisted real estate funds	Around 30%	Around 70%
Types of real estate assets held	Tend to be lower-risk or core real estate	Higher-risk private equity (PE) real estate funds
Life of fund	Indefinite life	Termination date, at which time the fund distributes all cash flows to investors
Liquidity (to investors)	Redemptions (monthly, quarterly, or annually); in the UK, PUTs also have a small amount of trading in the secondary market	Do not offer redemptions; investors can trade their positions in the secondary market (if it exists), although it is not liquid
Redemption price	Based on appraisals of properties owned by the fund	Based on proceeds from properties at the termination date
Fund structure	In the UK, PUTs are the most common open-end structure used by pension funds	Closed-end funds are very prevalent in the PE industry. In the U.S., most PE consists of limited partnerships or limited liability corporations; globally, limited partnerships are the most common structure
Tax status	Tax-free for qualifying pension funds	Limited partnerships are tax-neutral
Fees	50 to 200 bps (U.S.), and 20 to 150 bps (UK); yearly performance fees may also be charged	50 to 200 bps (U.S.), and 20 to 150 bps (UK); yearly performance fees may also be charged

2. Access to skilled managers (specialized in either a market sector or a specific geography) paves the way to higher returns and lower risk.
3. Some unlisted real estate funds offer the possibility to target subsectors or regions.
4. In general, unlisted real estate funds offer tax-exempt income to investors.

17.1.5 Three Disadvantages of Unlisted Real Estate Funds

Three disadvantages of unlisted real estate funds are:

1. Cash drag: Cash invested in unlisted real estate funds will not necessarily be drawn by the fund manager immediately. This occurs because cash will be drawn from investors as it is needed by the fund to purchase the respective real estate assets. Therefore, the investor will not be able to attain immediate full exposure to real estate assets when allocating funds to unlisted real estate funds.

2. Fees: Fees charged by unlisted real estate funds can be significant. Fees are charged yearly and amount to between 50 and 200 basis points in the United States and between 20 and 150 bps in the UK. Performance fees may also be charged (relative either to an index or to absolute returns). Performance fees can add up when the manager uses significant leverage and the appraisal values of the properties are increasing.
3. Leverage and the J-curve effect: The short-term performance of unlisted funds is usually affected by the costs involved in buying the initial portfolio of properties. Once fund costs are amortized, unlisted funds will have to outperform the direct market to compensate for the short-term underperformance. Leverage is one mechanism used to attain the desired outcome. It is important to note that while use of leverage is one of the advantages of real estate investment, it increases the risk of such assets, which could be exacerbated because the underlying asset is illiquid.

17.2 LISTED REAL ESTATE FUNDS

This section discusses real estate funds listed on exchanges, which include REITs, REOCs, and ETFs based on real estate indices. The section also presents the main advantages and disadvantages of publicly traded real estate funds. Finally, global REITs and non-traded REITs are discussed in the last part of this section.

17.2.1 REITs and REOCs

A real estate investment trust (REIT) is a professionally managed investment vehicle that invests in a portfolio of real estate assets, typically with the goal of generating income (mainly rental income) for its shareholders. REIT shares are listed on stock exchanges. In general, REITs do not have to pay corporate taxes on profits as long as a large proportion of income (90% in the United States) is distributed to shareholders as dividends. This feature, which is also present in the case of mutual funds, represents a major tax advantage when compared to the typical investment in stocks. Other important features of REITs were discussed in Level I.

Traditionally, REITs have been the most important listed real estate investment vehicle. REITs and analogous entities are common in the United States, the UK, Canada, Germany, Singapore, and Australia. These and many other countries have adopted the U.S. REIT approach. Also, in recent years, a number of alternative listed real estate investment vehicles have been launched. These new investment vehicles include options and futures on real estate indices, exchange-traded funds based on real estate indices, and closed-end real estate mutual funds.

Large REITs are actively managed and are vertically integrated firms, as they are involved in the following activities: land acquisition and holding, development, ownership (supply of financial capital and portfolio management), operation (asset and property management), and tenant services.

A **real estate operating company (REOC)** is similar to a REIT, except that a REOC reinvests its earnings into the business rather than passing them along to shareholders (and hence they do not get the same tax advantage enjoyed by investors in REITs). REOCs reinvest their earnings because their goal is to search for capital gains rather than passive cash flows. Furthermore, REOCs are more flexible than REITs in terms of the kinds of real estate investments they can make.

17.2.2 Exchange-Traded Funds Based on Real Estate Indices

As discussed earlier, most ETFs trade on exchanges at approximately the same price as the net asset value of their underlying assets due to provisions that allow for the creation and redemption of shares at NAV. ETFs have the advantage of being a relatively low-cost investment vehicle, are tax-efficient, and offer stocklike features (liquidity, dividends, the possibility to go short or to use with margin, and in some cases the availability of calls and puts on the ETF).

While the first ETFs were based on only stock and bond indices, ETFs are now also based on assets such as real estate, currencies, and commodities. Exchange-traded funds based on real estate indices and listed in U.S. exchanges track the following:

- Global real estate values (such as through the iShares S&P World Ex-U.S. Property Index Fund, the SPDR DJ Wilshire International Real Estate ETF, the WisdomTree International Real Estate Fund, and the iShares FTSE EPRA/NAREIT Global Real Estate Ex-U.S. Index Fund)
- Global regions (such as through the iShares FTSE EPRA/NAREIT Asia Index Fund, the iShares FTSE EPRA/NAREIT Europe Index Fund, and the iShares FTSE EPRA/NAREIT North America Index Fund)
- Broad U.S. exposures (such as through the First Trust S&P REIT Index Fund, the iShares Cohen & Steers Realty Majors Index Fund, the iShares Dow Jones U.S. Real Estate Index Fund, the iShares FTSE NAREIT Real Estate 50 Index, the streetTRACKS Wilshire REIT ETF, and the Vanguard REIT Vipers ETF)
- U.S. sectors (such as through the iShares FTSE NAREIT Residential Index, the iShares FTSE NAREIT Industrial/Office Index, the iShares FTSE NAREIT Retail Index, and the iShares FTSE NAREIT Mortgage REITs Index)
- U.S. mortgages (such as through the iShares GNMA Bond ETF, the iShares Barclays MBS Fixed-Rate Bond Fund, the SPDR Barclays Capital Mortgage Backed Bond ETF, and the Vanguard Mortgage-Backed Securities ETF)

There is even a real estate ETF with leverage (ProShares Ultra Real Estate) and an ETF with a short exposure (ProShares UltraShort Real Estate).[1]

Subject to the possible disconnect between private real estate values and public market prices, ETFs facilitate access to real estate assets for both small and large investors; and similar to options and futures on real estate indices, ETFs based on real estate indices can be used either to hedge risks or to speculate.

Finally, there are also ETFs that invest in REITs. For example, the iShares Cohen & Steers REIT (ICF) seeks to track the investment results of an index composed of relatively large and liquid U.S. REITs. Another ETF that invests in REITs is the iShares Real Estate 50 (FTY), which seeks to track the investment results of an index composed of the 50 largest real estate REITs within the FTSE NAREIT Composite Index. These two ETFs are the largest real estate ETFs by market capitalization in the United States (as of March 2015). Investors in ETFs that invest in REITs must bear two layers of fees: first from the REITs in which the ETF invests, and second from the ETF. Investors also bear brokerage fees when buying or selling ETFs in the market.

17.2.3 Six Advantages of Listed Real Estate Funds

Six advantages of listed real estate funds are:

1. Similar to unlisted real estate funds, listed real estate funds help diversify real estate specific risk. As noted previously, very large sums of money are needed to build a diversified portfolio of real estate physical assets. Listed real estate funds such as REITs facilitate exposure to investments in real estate assets requiring even lower sums of money than the typical open-end real estate fund.
2. Listed real estate funds have liquidity and divisibility. In general, REITs are fairly liquid investments. This is one of the main advantages of REITs compared to unlisted real estate funds. Also, their divisibility implies that investors can have indirect access to real estate assets in the precise amount that they wish to invest.
3. Listed real estate funds have instant exposure to a real estate portfolio. Cash invested in unlisted real estate funds will not necessarily be drawn by the fund manager immediately. This is because cash will be drawn from investors as it is needed by the unlisted fund to purchase the respective real estate assets. Therefore, the investor will not be able to attain immediate full exposure to real estate when allocating capital to unlisted real estate funds.
4. Listed real estate funds provide information to the investors. Similar to the case of other securities listed in exchanges, such as stocks and bonds, listed real estate funds are required to disclose financial and other related information about the fund. This allows investors to make relatively informed decisions.
5. Similar to unlisted real estate funds, some listed real estate funds offer the possibility to target subsectors or regions.
6. Similar to unlisted real estate funds, listed real estate funds offer some tax benefits, such as exemption from corporate taxes.

17.2.4 Two Disadvantages of Listed Real Estate Funds

Two disadvantages of listed real estate funds are:

1. REITs and even sometimes ETF shares may trade at a discount or premium to their NAV. This phenomenon has long been reported in the case of equity and bond closed-end mutual funds. In the case of U.S. REITs, there were fluctuations between a discount to NAV of 40% and a premium of 20% between 1990 and 2010. Several explanations have been offered to explain this phenomenon. For example, some attribute the problem to the practice of smoothing appraisals of direct investments made by REITs, which, they argue, render these valuations as less informative reflections of REIT values than REIT share prices. However, other market participants disregard the information content of REIT share prices because they contend that the prices are an unreliable indication of short-term market sentiment. This issue is discussed in more detail in later sections of this chapter.

EXHIBIT 17.2 Main Characteristics of Listed Real Estate Funds

	REITs	ETFs	Listed Funds and Mutual Funds
Types of real estate assets held	Can be equity REITs, mortgage REITs, or mixed	ETFs track a real estate index; there are also short and ultra-short ETFs	
Liquidity (to investors)	Traded on the secondary market	Traded on the secondary market	
Redemption price	Secondary market traded price	Secondary market traded price	
Tax status	REITs do not pay taxes at the company level if they comply with a number of restrictions; this means that no capital gains taxes are charged on the disposal of assets	ETFs are tax-neutral	Do not pay taxes at the company level if they comply with a number of restrictions; this means that no capital gains taxes are charged on the disposal of assets
Fees	Most REITs have annual managers' fees, property management fees, trustees' fees, and other expenses; brokerage fees must also be added when buying or selling REIT shares in the market	As most real estate ETFs invest in REITs, there are two layers of fees: the fees from the REIT and the fees from the ETF; brokerage fees must also be added	

The potential problem of ETF shares trading at a discount or premium to their NAV is a less severe problem than in the case of REITs. This is because most ETFs are constructed to facilitate arbitrage activities that drive the market price of the ETF toward the per-share value of the ETF's underlying portfolio, or its net asset value.

2. Listed real estate funds such as REITs may have a high correlation with the stock market (and not with the underlying real estate index). It has been observed that often the shares of REITs follow more closely the behavior of stock market indices than the trends of the real estate index underlying the REIT.

Exhibit 17.2 shows the main characteristics of REITs and ETFs regarding types of real estate assets that they hold, liquidity, redemption price, tax status, and fees.

17.2.5 Global REITs

Similar to the case of allocating funds to stocks on a worldwide basis, global real estate investing makes sense as long as the business cycles across countries are not

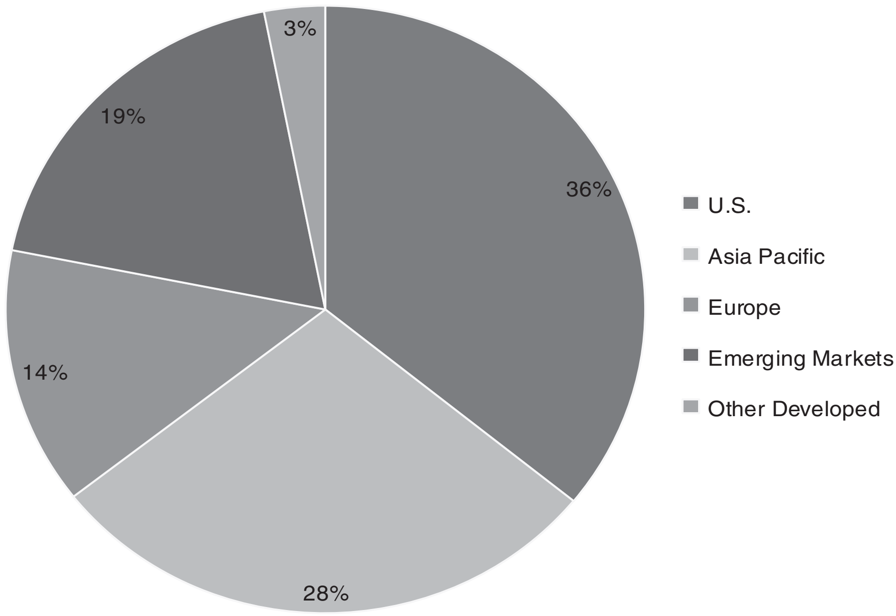


EXHIBIT 17.3 Size of the Global Real Estate Securities Market by Region, June 2014 (Total Size $1.72 Trillion)
Source: UBS and Cohen & Steers (www.cohenandsteers.com).

perfectly correlated. By investing internationally, an investor in real estate may well gain access to a higher efficient frontier (i.e., benefit from a higher expected return for a given level of risk) than if she had invested only in the domestic real estate market. We expand on this topic in the next chapter.

The global real estate securities market has a total market capitalization of $1.72 trillion (as of June 30, 2014), distributed across 456 companies. Exhibit 17.3 shows that the United States represents 36% of the total market, with 28% accounted for by the Asia-Pacific region and 14% by Europe. Emerging markets have witnessed an explosive growth in listed real estate in recent years, now accounting for 19% of the total market capitalization.

Global REITs are real estate investment trusts that invest in global real estate products. They are the premier listed investment vehicle for gaining access to international real estate markets. In general, and similar to the case with U.S. REITs, trusts on non-U.S. REITs do not have to pay corporate taxes on profits as long as a large proportion of income (90% in the United States, the UK, and Japan, and 85% in France, for example) is distributed to shareholders as dividends. Thus, the crucial tax-efficiency advantage of REITs is also present globally.

Global REITs also differ on the rules related to their management and their use of leverage. Many non-U.S. REITs are externally managed by outside companies. However, in the United States, REITs are normally managed by their executives or trustees. Regarding leverage, whereas U.S. REITs have no limits on the amount of debt financing they can employ, REITs in other countries do have debt ceilings. For example, Australian REITs have a debt-to-equity limit ratio of 3 to 1 if certain circumstances are met.

The various styles set forth by a REIT are decided by the fund management team, with the assistance of the CEO and the governing entity. The style can be active or passive; top down or bottom up; value or growth; and core, value added, or opportunistic. The style of REIT management is generally an active style, in which the manager actively runs an individual real estate investment or a portfolio of real estate assets, seeking returns above those realized by a benchmark within a certain level of risk. The style adopted by the investment management of REITs is usually top down. According to Parker (2011):

> *REITs have now evolved to focus primarily on the portfolio role of individual property assets with the top down style permeating the entire REIT real estate investment decision making process through the steps of strategic asset allocation (wherein the entire real estate market is analyzed through a range of risk-return lenses to identify target sectoral and geographic weightings), tactical asset allocation (wherein target markets offering significant potential outperformance are identified) and stock selection (being the process of identifying portfolio asset criteria).*

Regarding the value versus growth dilemma, REIT management is normally a growth style, though tactical asset allocation may incorporate components of a value style. According to Parker (2011), the crucial question is the specific style taken by a REIT: that is, core, value added, or opportunistic.

The global REIT market has experienced rapid growth during the past decade, particularly in the case of REITs from emerging markets such as China and India. Approximately three-quarters of the companies in the global real estate securities market consist of REITs or REIT-like structures, with the remaining portion accounted for by non-REIT owner-operators and real estate development companies. Exhibit 17.4 shows the countries with listed REITs, as well as the countries for which REIT legislation is either in progress or under consideration. The first country that adopted the REIT structure was the United States (1960), followed by the Netherlands (1969) and Australia (1971). The next country was Canada, which adopted REITs in 1994, 23 years later. The remaining countries introduced REITs in the years that followed, meaning that for many countries, REITs have been in existence for less than two decades.

Exhibit 17.5 shows the average 10-year return and standard deviation of REITs from nine developed countries, the only countries for which a 10-year series was available. It can be observed that the range of standard deviations is relatively narrow (from Belgian REITs, the lowest at 17.57%, to Singaporean REITs, the highest at 26.20%). However, the range of average yearly returns is wider, from a low of 5.10% for Australia to a high of 16.46% for France. The average of the REIT returns from these nine countries (equally weighting each country) is 10.60%, which is only slightly higher than the U.S. average. These are respectable returns, considering that the global real estate and financial crises took place during this period.

Exhibit 17.6 shows the evolution of $100 invested in global REITs (measured by S&P Global REITs Total Return Index) compared to the evolution of $100 invested in three other types of global investments (world equities, measured by MSCI World TR Equities; global bonds, measured by J.P. Morgan Global Aggregate; and global commodities, measured by the S&P GSCI). The graph begins in March 2005 because

EXHIBIT 17.4 Countries with Listed REITs

Countries with Listed REITs and Year Adopted		REIT Legislation in Progress	REIT Legislation under Consideration
United States (1960)	Malaysia (2005)	Chile	China
Netherlands (1969)	Israel (2006)	Costa Rica	India
Australia (1971)	Germany (2007)	Hungary	
Canada (1994)	UK (2007)	Indonesia	
Ghana (1994)	Italy (2007)	Lithuania	
Belgium (1995)	New Zealand (2007)	Luxembourg	
Brazil (1995)	Nigeria (2010)	Philippines	
Greece (1999)	Mexico (2011)	Puerto Rico	
Turkey (1999)	Thailand (2012)		
Japan (2000)	Finland (2013)		
South Korea (2001)	Ireland (2013)		
Singapore (2002)	Pakistan (2013)		
France (2003)	South Africa (2013)		
Hong Kong (2003)	Dubai (2014)		
Taiwan (2003)	Spain (2014)		
Bulgaria (2005)			

Source: UBS and Cohen & Steers (www.cohenandsteers.com).

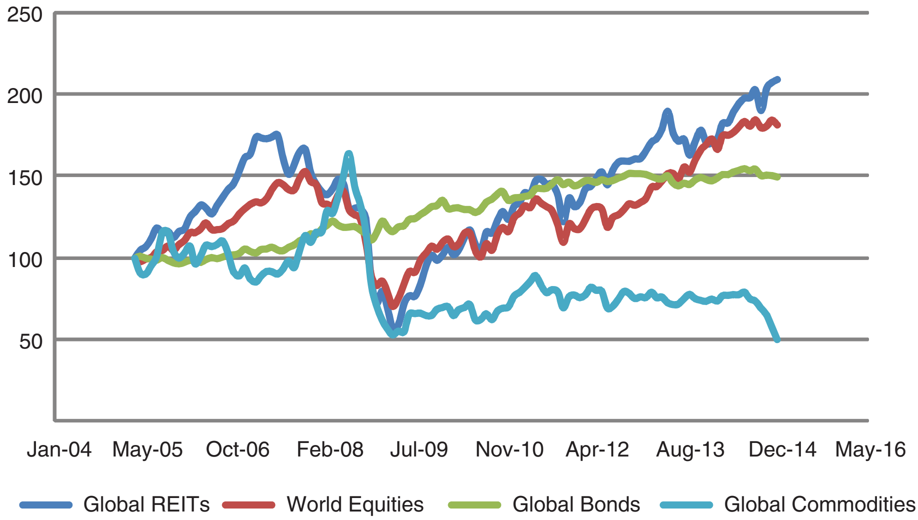
this is the first date for which S&P reported the value of its then newly created Global REITs Index. We can observe that the evolution of global REITs is highly correlated to the evolution of world equities (the correlation coefficient is 0.84). Global REITs experienced average monthly returns and standard deviation of 0.88% and 6.64%, respectively, which were slightly higher than those of world equities, which were 0.62% and 4.65%, respectively. And similar to world equities and global commodities, but unlike global bonds, global REITs suffered extensively as a result of the global financial and real estate crisis of 2008–9, beginning to decline in 2007 and

EXHIBIT 17.5 Ten-Year REIT Index Performance for Nine Developed Countries (Up to June 2014)

	Average Annual Return	Annual Standard Deviation
Australia	5.10%	25.16%
Belgium	7.36%	17.57%
Canada	14.83%	20.68%
France	16.46%	25.64%
Japan	7.78%	20.91%
Netherlands	8.09%	24.55%
New Zealand	11.07%	20.68%
Singapore	15.09%	26.20%
United States	9.58%	25.79%

Source: Dow Jones Real Estate Indices, S&P Global Property Indices, quantitative analysis, Q2, 2014.

EXHIBIT 17.6 Cumulative Wealth Index: Global REITs versus Other Investments, 3/2005–2014
Source: Standard & Poor's and Bloomberg.



experiencing a maximum drawdown of close to 70% from that moment until early 2009.

The high correlation exhibited between global REITs and world equities highlights once again one of the major drawbacks related to REIT investing: the fact that REITs exhibit a high correlation with stocks. This is discussed in greater depth later on in the chapter.

Finally, there are ETFs that invest in the international real estate market. For example, the iShares International Developed Property ETF seeks to track the investment results of an index composed of real estate equities (REITs) in developed non-U.S. markets. Once again, investors in ETFs that invest in REITs must bear two layers of fees (first from the REITs in which the ETF invests, and second from the ETF).

17.2.6 Non-Traded REITs

Although non-traded REITs are not listed on any exchange and should therefore not be considered listed REITs, they have some features in common with listed REITs. For instance, in the United States, they are registered and available to retail investors. **Non-traded REITs** were created in the United States in 1990, and even though they are registered with the Securities and Exchange Commission as public companies, their shares are not available on an exchange and are thus essentially illiquid (and therefore difficult to value). Investors can usually sell their shares after a year and under a limited repurchase program.

In the United States, shares on non-traded REITs are available through registered financial advisers. After the financial crisis, non-traded REITs gained popularity among retail investors because of their relatively large dividends, which compared favorably to common stock dividends and bond yields.

The estimated life span of this type of REIT is usually seven to 10 years, at which time it ends in a liquidity event. Just like the case of traded REITs, non-traded REITs must pay out at least 90% of their taxable income to shareholders in order to preserve their status as being exempt from federal corporate income taxes. Non-traded REITs are available only to investors who meet suitability standards, and they represent about 20% of the total market capitalization for all publicly registered REITs (in the United States).

Up-front fees associated with investing in a non-traded REIT are usually between 12% and 15%. One of the major challenges for non-traded REITs occurs during the ramp-up phase, when investors have the expectation of perceiving future dividend payments at a time when the REIT needs to cover relatively high initial fees and there are only a few assets ready to generate cash flow.

The life cycle of a non-traded REIT has four distinct phases, although in practice some of the phases may overlap to some extent. The first phase is the capital-raising stage. Usually no additional new capital is available after this initial stage ends. In the second phase, the REIT acquires its portfolio of properties with the capital raised in the initial phase. In the third phase, the REIT manages the assets it owns, attempting to generate positive cash flows and to increase value (this is the asset management phase). The final stage is known as the disposition phase. During this stage, an exit strategy is executed to return the investors' original investment and any capital gains or losses that may result from the liquidity event.

Similar to the case of traditional REITs, non-traded REITs can be classified into three main categories: equity, mortgage, and hybrid. Furthermore, there are seven broad property categories in which a traded or non-traded REIT may invest: (1) apartments or multifamily properties, (2) office space, (3) industrial facilities, (4) retail space, (5) hospitality properties, (6) health-care properties, and (7) self-storage properties.

Non-traded REITs can explicitly select two types of strategies or a combination of the two. The first is a current income strategy (also known as distributions), which is associated with higher current yields but at the expense of lower overall total returns. The second is a long-term price appreciation strategy. In the case of a non-traded REIT, this strategy is realized at the sale of the assets.

Non-traded REITs have received the following three main criticisms (Husson, McCann, and Taveras 2012). First, their illiquid nature may give investors a misleading sense of low return volatility. Second, and as previously noted, they command high fees, and often involve significant conflicts of interests. Third, they often use leverage to fund current dividend payments, a practice that may hide their incapability to generate future dividends.

Finally, Exhibit 17.7 briefly compares traded and non-traded REITs as they pertain to investment objective, liquidity, volatility, investors, and fees.

17.3 MARKET-BASED VERSUS APPRAISAL-BASED RETURNS

Chapter 15 provided an analysis of the differences between market-based and appraisal-based real estate returns over an eight-year period that included the global financial crisis that began in 2007. The analysis of Chapter 15 discussed the delayed

EXHIBIT 17.7 Comparison between Traded and Non-traded REITs

	Traded REITs	Non-Traded REITs
Investment objective	Current income and share price appreciation	Current income and eventual capital appreciation
Liquidity	May be bought and sold at any time; either listed on an exchange or traded over-the-counter (OTC)	Investors should plan to hold through planned exit event
Volatility	Daily share price volatility based on supply and demand	No daily share price volatility until repricing 18 months after close of offering period, when underlying value of real estate holdings may fluctuate up or down
Investors	Available to any person or entity with a brokerage account	Available only to qualified investors through a professional financial adviser
Fees	Trading commissions at purchase; annual asset-based management fees	Front-end fees of 12% to 15% of investment; acquisition, asset management, disposition, and incentive fees during term

Source: Taken from Investment Program Association (2013).

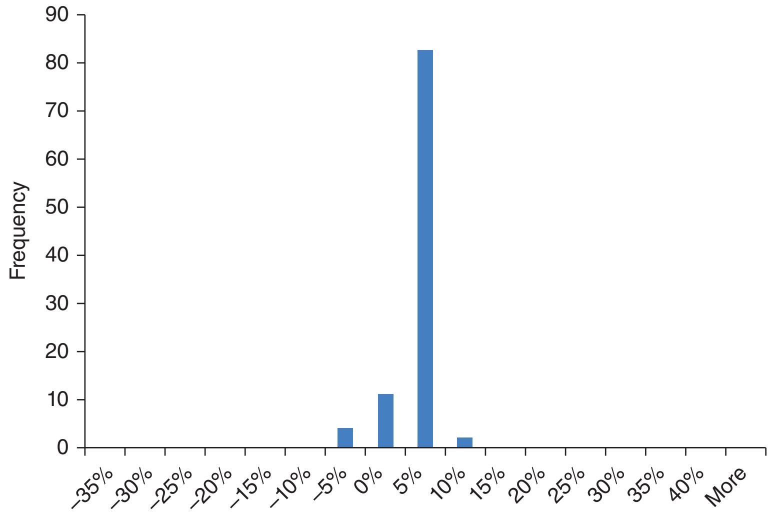
reaction of appraisal-based returns relative to market-based returns, and illustrated unsmoothing. This section analyzes data on appraisal-based and market-based U.S. real estate returns over a 25-year period.

17.3.1 Histograms of U.S. Real Estate Returns

Exhibit 17.8 is a histogram of the quarterly returns of the NCREIF Property Index (NPI) over the 25-year period from the first quarter of 1990 through the fourth quarter of 2014. The simple average quarterly return of the 100-quarter return distribution is 1.9%, and the quarterly standard deviation is 2.4%. The critical issue is whether the concentration of outcomes near the center in Exhibit 17.8 is an accurate reflection of the stability of real estate values or a result of smoothing due to the appraisal methodology.

Next, market-based returns are examined as a contrast to appraisal-based returns. Exhibit 17.9 is a histogram of the quarterly returns of the All Equity REITs Index of the FTSE NAREIT US Real Estate Index Series (REIT Index) over the same 25-year period from 1990 through 2014. The REIT Index has a simple average quarterly return of 2.9% and a quarterly standard deviation of 9.9%. Note that the annualized standard deviation of 9.9% quarterly (assuming no autocorrelation) is 19.8%, which is in line with the very long-term annualized standard deviation of value-weighted indices of U.S. equity returns.

EXHIBIT 17.8 NCREIF Property Index (NPI)

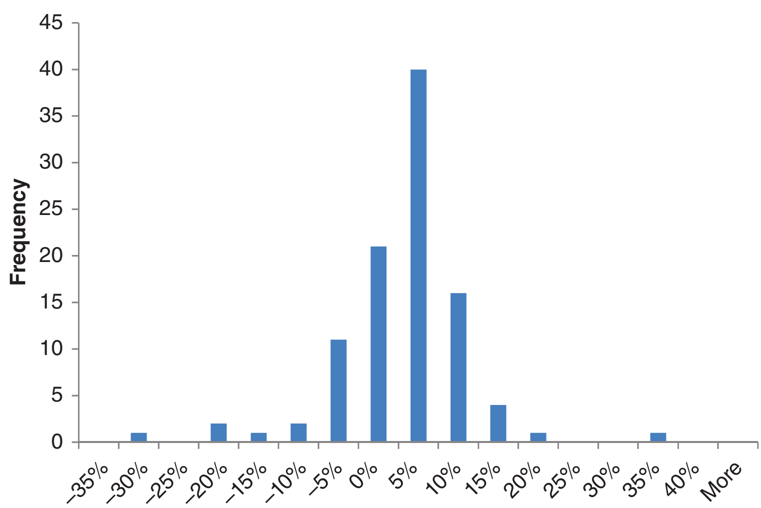


17.3.2 Four Explanations of the Relationship of Appraisal-Based and Market-Based Volatilities

The critical issue is whether the much wider dispersion of the market-based returns (Exhibit 17.9) relative to the appraisal-based returns (Exhibit 17.8) is due to one or more of the following four explanations:

1. Lower accuracy of smoothed appraisals relative to market prices
2. Differences in the risks of the underlying assets

EXHIBIT 17.9 REIT Returns



3. The use of leverage in most REITs
4. The result of liquidity-induced volatility attributable to liquidity shifts in equity markets

Regarding the lower accuracy of appraisal-based pricing, there is evidence regarding smoothing. Cannon and Cole (2011) compared appraised real estate values from the NPI over a 25-year period with transaction prices on the same properties observed within two quarters of the appraisals. They report: "Our findings are sobering. On average, appraisals are more than 12% above or below subsequent sales prices, and this result holds true for both external and internal appraisals." In attributing the price differences, they find that "appraisals appear to lag the true sales prices" and that "under- and overvaluations are highly correlated across properties at the same points in time." This means pricing errors are systematic and are correlated on a cross-sectional basis.

In the first of these explanations, REIT returns are viewed as superior indications of real estate risks and returns when compared to appraisal-based returns, because market-based returns reflect valuations from equity markets with a moderate to high degree of informational market efficiency. To the extent that the market prices of REITs differ substantially from their true values, it can be argued that market participants would be able to earn superior rates of return by purchasing underpriced REITs or short-selling overpriced REITs. Presumably, competition to trade mispriced REITs will tend to drive REIT prices toward their true values. Appraisals are not generally subject to arbitrage and therefore may be more likely to diverge from the true values of the underlying property.

The second potential explanation of the observed differences between REIT returns and appraisal-based returns is that the two return series are based on substantially different underlying assets. However, most of the equity REITs underlying the REIT Index analyzed in Exhibit 17.9 own portfolios of large institutional-quality properties highly similar to the core properties that dominate the NPI.

The third possible explanation is that leverage accounts for the higher volatility of the market-based index (the REIT Index) relative to the appraisal-based index (the NPI). The REIT Index returns include the effects of leverage, but the NPI is an unlevered index. Although much of the property that underlies the NCREIF Property Index is held on a leveraged basis, when its financial performance is reported, it ignores financing costs and uses asset values rather than equity values. In the real estate industry, leverage is a key component of most property purchases. Almost all institutional investors use leverage of perhaps 50% in the purchase of their real estate portfolios, which should result in a doubling of equity volatility relative to asset volatility. Equity REITs use leverage averaging roughly 40%. However, returns of the market-based REIT Index exhibited four times the volatility of the appraisal-based NPI returns. Leverage differentials can probably explain less than half the difference in volatility.

The fourth potential explanation is that equity market trading of real estate investments injects volatility into the REIT prices, which is attributable to equity market volatility and liquidity rather than to true changes in the values of the underlying real estate properties.

In summary, even adjusted for leverage, appraisal-based returns have dramatically lower volatility than do market-based returns. Chapter 15 provided a brief analysis of the degree of smoothing in the NCREIF Property Index. The evidence of smoothing in appraisal-based indices is strong, and therefore the first of the four potential explanations (lower accuracy of smoothed appraisals relative to market prices) may account for all of the leverage-adjusted volatility differences. But the potential remains that the divergence between the NPI and the REIT Index is attributable at least in part to inefficient pricing of REITs during periods of high equity market volatility. This issue, revisited later in the chapter, is critical in determining whether market prices or appraisals should be used in estimating real estate risks.

17.3.3 The Importance of Accurate Pricing and Risk Estimation

The previous sections illustrated the large volatility differences between appraisal-based and market-based real estate index volatilities and discussed potential explanations. The two valuation approaches offer substantially different indications of the volatilities and correlations of real estate returns. The importance of accurate risk estimation can be viewed at two levels.

On an investor-based level, better portfolio management decisions can be made with better estimates of risk. To the extent that investors are using appraisal-based returns to form estimates of real estate risk, they will require an insufficiently low return on the real estate if the returns understate risk. As a result, the portfolio will be overallocated to real estate, and the portfolio manager will be underestimating the portfolio's level of risk.

From a macroeconomic perspective, accurate pricing of assets requires accurate estimation of risk. If real estate investors as a group underestimate the required return on real estate investment due to underestimation of real estate risk, the asset class will be overpriced, and the level of investment made in real estate will be inefficiently high. In other words, to the extent that real estate risk is underestimated, real estate projects offering relatively low rates of return (i.e., relatively low benefits to society) will be funded even though the risks of those projects, if correctly estimated, would indicate that the projects reduce wealth relative to other investments.

17.4 ARBITRAGE, LIQUIDITY, AND SEGMENTATION

The divergence between the returns of appraisal-based real estate indices and those of market-based real estate indices should not be lightly dismissed as being fully attributable to the smoothing caused by appraisals. Appraisal methodologies use the same information and follow the same decision-making process as investors who trade private commercial real estate. Especially in times of extreme market stress, investment professionals within the real estate industry are likely to place less reliance on REIT values than on fresh appraisals as indications of the true value of real estate.

This section discusses explanations for the divergence between appraisal-based indices and market-based indices in the context of similar circumstances with other investments. Arbitrage drives similar assets toward similar prices. Real estate professionals should view the tendencies of similar assets to converge and diverge in value in investments other than real estate to develop a clearer understanding of real estate prices.

For example, the availability of securities offering highly liquid exposure to asset classes and asset subclasses within traditional investments plays an important role in traditional portfolio management. Exchange-traded funds are an example of an innovation that has tremendously facilitated risk management for some assets. Do similar tools exist within real estate, or are they emerging? This section begins with a discussion of ETFs in order to introduce related opportunities and challenges in real estate risk management.

17.4.1 Pooling of Securities versus Securitization

An important distinction exists between the pooling of securities and securitization. The **pooling of securities** is any collection of securities in a single entity, such as a mutual fund. Securitization is generally defined as the pooling of non–publicly traded assets into publicly traded ownership units (securities). Thus, an ordinary equity mutual fund is generally viewed as a pooling of securities but not as a securitization, because the underlying equities are already publicly traded securities.

Publicly traded real estate funds, such as most REITs, are forms of securitization, since they hold underlying assets that are not publicly traded securities. ETFs generally have publicly traded securities as their underlying assets. As indicated in the next few sections, the liquidity of the underlying assets has important implications for the relationship between the pool's value and risk and the value and risk of the underlying securities.

17.4.2 Exchange-Traded Funds and Arbitrage

The emergence of ETFs has substantially improved the ability of traditional equity portfolio managers to control their risk exposures quickly and cost-effectively. What makes ETFs so popular and useful? The answer is not just that the units themselves are tradable. Many types of securities exist, such as closed-end mutual funds, that are themselves liquid and that have publicly traded underlying assets. Nevertheless, many closed-end mutual funds have limited usefulness for hedging and benchmarking. The difference between closed-end mutual funds and ETFs is that the relationship between the closed-end funds and their underlying securities is not easily arbitraged.

The key to ETFs is the ability of market participants to arbitrage any difference between the market price of the ETF and the per-share value of its underlying portfolio. Most ETFs are constructed to facilitate arbitrage activities that drive the market prices of the ETF toward the per-share value of the ETF's underlying portfolio, or its net asset value.

For example, most ETFs allow owners of shares in the ETF to redeem their ETF shares for shares in the portfolio of securities that underlie the ETF. If an ETF trades at a substantial discount to its NAV, an arbitrageur can (1) buy the ETF shares in the market, (2) tender the shares to the corresponding fund in exchange for shares in the ETF's underlying portfolio, and (3) sell the shares received from the tender into the market at a net profit. The extent to which the ETF's market price is below its NAV will determine the profit of this arbitrage transaction before transaction costs. Thus, ETF prices should not stray from underlying NAVs by more than the transaction costs involved (provided the ETF's NAV is based on market prices).

Conversely, most ETF trading is performed with ETFs that allow large investors to tender the securities underlying the ETF to the fund in exchange for new shares in the ETF. If an ETF becomes overpriced relative to its NAV, an arbitrageur can (1) short-sell the ETF in the market, (2) buy eligible and representative shares for the ETF's underlying portfolio, (3) convey the shares to the corresponding fund in exchange for the acquisition of new shares in the ETF, and (4) use the new ETF shares to cover the original short position in the ETF. This arbitrage activity of short-selling ETF shares will drive down the price of the ETF until any excess of the ETF's market price above its NAV is at least as small as the transaction costs involved.

The key benefit of the potential of many ETF prices to be arbitraged when they deviate substantially from their NAVs is that market participants can establish long and short positions in ETFs with confidence that ETF prices will be near their NAVs not only during periods of market calm but also during periods of moderate stress. This liquidity means that the risk of traditional equity portfolios can be continuously and effectively managed using ETFs. In other words, traditional equity portfolio management can rely on efficient pricing of highly liquid funds such as ETFs, options, and futures contracts to manage risks. Should real estate investors similarly rely on publicly traded real estate funds such as REITs to manage the risks of private real estate equity holdings?

17.4.3 The Hedging of Unlisted Real Estate Values with Listed Real Estate Values

Consider an institution wishing to hedge some or all of its position in private real estate equity. The position consists of highly illiquid holdings in limited partnerships corresponding to a variety of U.S.-based private properties. The institution is considering hedging its real estate holdings with an ETF product that correlates to a broadly diversified major index reflecting the overall performance of equity REITs in the United States. The institution considering the hedging notes that the properties underlying the REITs and the properties underlying the institution's private property holdings are highly similar in terms of type, size, and general location. Given the similarity of the underlying real estate and substantial evidence regarding the level of informational efficiency of major financial markets, can an ETF of REITs serve as an effective hedge of true risks?

The evidence in Chapter 15 and in Exhibits 17.8 and 17.9 indicates that publicly traded real estate securities exhibit extremely high volatility, even adjusted for leverage differentials, when compared to appraisals of private real estate. In periods of market stress, the divergence can be startling. For example, Exhibit 15.2 in Chapter 15 lists the quarterly returns of the NPI and the REIT Index for the eight years ending in December 2014. Note that beginning in the fourth quarter of 2008, the REIT Index declined approximately 60% in value within six months. However, in the next nine months, the REIT Index roughly doubled. Thus, a \$100 position in the index would have declined to \$41 and recovered to \$78 in just five quarters. During the first six months of the same period, the NPI fell only about 15%, but then continued falling another 8% in the next six months, rather than rebounding like the REIT Index did.

The key question to the institution considering a hedge for its private real estate holdings is this: Did the REIT Index's movement correlate highly with true private real estate values? If so, the REIT Index will serve as a suitable hedge for its private positions even if appraised prices indicate otherwise. If not, then the REIT Index may be adding unnecessary risk.

In this example, the REIT Index clearly did not move in tandem with the NCREIF appraisal-based index. To the extent that the underlying properties are similar, one or both of the indices must be in substantial error. There are reasons to believe that appraisal-based valuations contain substantial errors. Thus, a reasonable argument exists that private real estate portfolio managers should rely on the informational efficiency of publicly traded REIT prices, and use REITs and REIT ETFs as risk management tools.

But publicly traded REIT valuations could also be in substantial error, in which case REITs would serve as poor risk management tools. The reason that REIT prices may substantially diverge from reflecting the true values of their underlying properties is that there is no method of quickly and safely arbitraging perceived differences between REIT valuations and their underlying assets.

17.4.4 Two Views of REITs as Indicators of Private Real Estate Values

The previous example of the performance of the REIT Index relative to that of the NCREIF Property Index in 2008 and 2009 indicated tremendous divergence in valuation. There are two primary interpretations of this divergence in performance:

1. The REIT Index's substantial price decline followed by a near 100% price increase accurately represents the true changes in the values of real estate properties adjusted for the effects of leverage.
2. The REIT Index's returns and their high volatility emanate from a source of risk uncorrelated with the underlying economic fundamentals of the real estate, such as volatility in the U.S. stock market driven by illiquidity and stress.

In the first explanation, REIT prices are informationally efficient. In the second explanation, REIT prices can diverge enormously from reflecting true real estate values due to market stress. There is substantial disagreement on which of these explanations is more accurate.

If the first explanation is accurate, then relevant REIT Index values and the true value of private real estate are highly correlated. Accordingly, market-traded real estate funds can be used to hedge the risks of private real estate. If the volatility levels between REITs and an investor's private properties differ due to leverage, the hedge ratio can be adjusted and an effective hedge can be maintained as long as the two real estate values (true private values versus public prices) are highly correlated.

But if the second explanation is accurate (that REIT-based returns are poorly correlated with true underlying private real estate values), then the hedge will tend to be ineffective, especially in times of equity market turbulence. In times of high equity market volatility, the hedge may actually add to the total risk of the investor.

The problem with resolving whether publicly traded real estate values are highly correlated with private real estate values is that, unlike with publicly traded securities, values of private property cannot be observed unless they are traded frequently.

Empirical efforts to resolve the issue are not definitive. However, Fisher et al. (2003) empirically analyzed publicly traded real estate versus appraised private properties and concluded that "the general pattern of price discovery seems to involve the NAREIT Index typically moving first . . . followed last by the appraisal-based NCREIF Index. The total time lag between NAREIT and NCREIF can be several years, as measured by the timing of the major cycle turning points."

17.4.5 Financial Market Segmentation

In financial markets, market segmentation refers to the differences in pricing of similar assets trading in separate markets, attributable to the differences in the markets themselves. These market differences are typically attributed to the differences in the clienteles that participate in the markets. In finance, a **market clientele** is a general type of market participant that dominates a particular market.

For example, two futures contracts may trade in different markets but with the same underlying assets. One of the futures contracts will have a large notional value and be traded on an exchange dominated by institutions and other large investors. The other futures contract will have a much smaller notional value and be traded on an exchange dominated by much smaller investors. Market segmentation in this instance would refer to differences in price, risk, and returns of the two contracts, attributable to the differences in the clienteles of the market. Depending on the circumstances, arbitrageurs may or may not be able to integrate the two markets.

The evidence in this chapter that REIT valuation and private real estate valuation diverge substantially may indicate that the real estate investment market is segmented. Many institutions, such as endowments, may have limited need for liquidity throughout their portfolios and may view their investments with longer-term horizons. These investors may be drawn to private real estate investments in the belief that these investments have relatively low risk, given their steady cash flows and relatively low dispersion in values at the longer-term horizon point at which the investment might be liquidated. Such institutions may view REITs as riskier due to their volatile market prices.

Investors with shorter-term horizons and a higher need for liquidity may perceive REITs as providing the liquidity they desire with low transaction costs. These shorter-term investors may perceive private real estate as having just as much risk as or higher risk than REITs because of the steep price discounts and high transaction costs that they may be forced to bear in liquidating private real estate during periods of stress.

Within this market segmentation view, both public and private real estate are accurately priced for their respective clienteles. Barriers to arbitrage activity, such as transaction costs and short-selling limitations, allow the valuations to diverge on a short-term basis.

17.4.6 Real Estate Turnover, Dealer Sales, and Agency Costs

It is possible that the legal and managerial structure of REITs may cause REIT market values to diverge from the values of privately owned real estate. For example, Mühlhofer (2013) provides a tax-based rationale for the apparent divergence between U.S. REIT market prices and appraisal-based indications of private real estate prices based on rules regarding dealer sales.

Dealer sales are transactions that are taxed unfavorably in the United States based on indications that the investor is a shorter-term dealer in the assets being sold rather than a long-term investor. The rules regarding dealer sales mean that if a REIT trades a particular percentage of properties within a particular time period and with a particular holding period, any capital gains from the transactions will be fully taxed at the REIT level.

Mühlhofer (2013) argues that REIT managers are constrained by the tax rules into suboptimal portfolio management, including reluctance to exploit inefficient private real estate values. The study examines the relationship between returns from the REIT market and those from the direct property market. REITs face the dealer rule, which prohibits them from purchasing real estate for sole purpose of reselling it. This indicates that REITs are property income vehicles, rather than complete property investment vehicles. A REIT investor has lower exposure to the short-term growth in the portfolio's assets while having a significant exposure to the rental returns from the REIT's portfolio.

In 2008, the rules regarding dealer sales were relaxed. An act of the U.S. Congress revised the taxability of REITs engaged in turnover (sales) of real estate properties. The legislation enabled REITs to buy and sell assets more effectively. As already mentioned, a REIT does not pay taxes at the corporate level, and to maintain REIT status, it must meet certain income and asset tests, which ensure that a REIT generates most its income in a passive way from real estate related income. In addition, a REIT is subject to higher taxes on the income from the sale of property held primarily to sell. The taxes can be avoided if the property is held for four years and the sales are do not exceed 10% of the REIT's holdings. The new legislation reduces the holding period requirement from four years to two years and relaxes some aspects of the 10% limitation (Edwards and Bernstein 2008; Humphreys 2008).

Given the relaxation of the rules regarding dealer sales, the evidence that REIT values and appraisal-based values diverged substantially through 2008 and 2009 (as discussed earlier) undermines the argument that taxation of REIT turnover causes these valuation differences. However, differences in agency relationships may also cause differences between the values of unlisted and listed real estate. For example, REITs have managers with dispersed shareholder bases, whereas most major real estate partnerships are formed and monitored by relatively small groups of limited partners. The reduced agency costs and higher responsiveness of private real estate managers to institutional investors may lead private real estate managers to pursue lower-volatility strategies. Conversely, REIT managers can be highly responsive to shareholders through such incentives as stock options.

17.4.7 Real Estate Price Volatility and Liquidity

Advocates of the informational efficiency of major securities markets, such as the markets in which U.S. REITs trade, argue that market prices of REITs provide better indications of value than do appraised prices of real estate. Within this view, REIT returns reflect true (but leveraged) changes in the economic value of the underlying real estate. In other words, when REIT returns indicated substantial and quick declines followed by a partial rebound, the values were accurate indications of the prices at which typical private commercial real estate was being transacted (adjusted for leverage).

Did private commercial real estate properties trade at tremendously reduced values in the spring of 2009, and did those values substantially rebound by September 2009? More generally, do actual private property transactions fluctuate in line with the value changes reflected in public REIT prices?

Practitioners who liquidated U.S. property holdings under duress in late 2008 and early 2009 may have had to suffer the same low prices implied by REIT prices due to the financial and credit crisis of the time. To real estate investors with liquidity problems, the significant decline in real estate prices may have been a valid indication of the prices they would be forced to accept in liquidating existing positions. But to well-capitalized real estate investors, it may be argued that the fire-sale prices were not accurate indications of the true value of their holdings. A foundation of this argument is that during a liquidity crisis, there is a very high likelihood that real estate prices will rebound when the crisis eases. If so, it stands to reason that the value of private real estate holdings of an investor certain to be able to ride out the crisis without liquidating positions might have been better indicated by professional appraisals than by market-based indications of value.

The hypothesis that short-term REIT prices imply erroneous indications of the value of a portfolio of real estate that is well capitalized suggests that publicly traded real estate securities do not offer effective risk management tools for portfolio management of illiquid assets even when the underlying assets are similar.

In the case of holdings of publicly traded assets, the actions of arbitrageurs help force ETFs and other products, such as derivatives, to maintain pricing relationships that facilitate cost-effective hedging. But alternative investments are often illiquid. Therein lie the challenges of investment management, as well as the potential rewards of alternative investments from the illiquidity and complexity premiums that may be available.

In the case of private real estate, if short-term REIT prices are at times truly out of line with the value of the underlying well-capitalized office and industrial properties, then arbitrageurs could presumably exploit this divergence by purchasing the underpriced asset and short-selling the overpriced asset. But there are two difficulties with this strategy in the case of REITs and real estate properties. First, the purchase and subsequent sale of private real estate take substantial time, transaction costs, capital, and expertise. Second, the short sale of privately held real estate is virtually impossible, and even the short sale of a REIT may be difficult in times of extreme market stress.

17.4.8 Evidence from Correlations over 20 Years

Exhibit 17.10 plots the 20-quarter rolling correlation between the quarterly returns of the appraisal-based NPI and the concurrent quarterly returns of the market-based all-equity sub-index of the FTSE NAREIT US Real Estate Index Series. Each point in the exhibit is based on the 20 quarters of returns leading up to the date on the horizontal axis. Casual observation of the diagram indicates virtually no consistently positive or negative correlation between the returns of the two real estate indices over the entire time period. The general range of the correlation is between –0.30 and +0.30.

The rolling correlation does spike up at the end of 2008, when two quarters of simultaneous large negative returns registered as the financial crisis that began

EXHIBIT 17.10 Rolling Correlation between NAREIT and NCREIF Index Appraisals

Source: Bloomberg and author's calculations.

in 2007 deepened. These returns were shown in detail in Exhibit 15.2 in Chapter 15. Note, however, that the correlation between the two indices in Exhibit 17.10 plummets back within its historical range beginning in 2009 (except for a brief period in mid-2014), when the indices move in opposite directions. (As shown in Exhibit 15.2 and discussed previously, the REIT Index recovered while the NPI continued to drift downward.)

The evidence from Exhibit 17.10 shows that other than during periods of extraordinary market stress, there is no historical basis to indicate that REITs can serve as an effective short- to intermediate-term hedging vehicle for appraised values.

17.4.9 Real Estate as a Diversifier

Exhibit 17.11 contains summary statistics for U.S. equity market indices (Russell indices), the NCREIF Property Index, the REIT Index, and unsmoothed versions of the NCREIF Property Index and the REIT Index. The analysis covers the 24.75-year period from the second quarter of 1990 through the first quarter of 2015. The exact time interval was based on data availability of the Russell indices. The unsmoothed returns for the NCREIF Property Index and the REIT Index were based on the first-order autocorrelation method detailed in Chapter 15.

First, note from Exhibit 17.11 that the correlation of the NPI with the REIT Index is only 0.17. In finance, prices of assets tend to exhibit higher correlation when returns are measured in longer intervals of time (monthly and quarterly) compared to shorter intervals of time (minutes and days). The reason for this is that returns over very short periods of time tend to be dominated more by idiosyncratic returns than by systematic returns. Thus, a correlation of only 0.17 between quarterly returns of indices with highly similar assets raises serious questions about the reliability and accuracy of one or both of the indices. It should be noted that the leverage implicit in

EXHIBIT 17.11 Correlations of Returns on U.S. Real Estate and Equity Indices

	NPI	REIT Index	Unsmoothed NPI	Unsmoothed REIT Index
NPI	1.00	0.17	0.59	0.15
REIT Index	0.17	1.00	0.42	0.99
Unsmoothed NPI	0.59	0.42	1.00	0.41
Unsmoothed REIT Index	0.15	0.99	0.41	1.00
Russell 3000	0.12	0.61	0.23	0.62
Russell Midcap	0.09	0.67	0.26	0.68
Russell Midcap Value	0.12	0.78	0.32	0.78
Russell Midcap Growth	0.06	0.48	0.17	0.50
Russell 2000	0.08	0.69	0.26	0.70
Russell 2000 Value	0.10	0.80	0.32	0.79
Russell 2000 Growth	0.06	0.55	0.19	0.57

the REIT Index will cause much higher volatility in that index relative to an unleveraged index (the NPI). But leverage should not have a major effect on correlations. For example, consider two portfolios with almost identical underlying assets, except that portfolio L is leveraged 2:1 and portfolio U is unleveraged. If the leverage ratio is continuously maintained, the returns of portfolio L should have twice the volatility of the returns of portfolio U but a correlation coefficient equal to 1.0.

Second, note in Exhibit 17.11 that the correlation of the NPI with the most general of the U.S. market indices listed (the Russell 3000) is 0.12, while the correlation of the REIT Index with the Russell 3000 is 0.61. Also note, however, that the NPI does not have a higher correlation with any of the other Russell equity indices, while the REIT Index has its greatest correlation (0.80) with the Russell 2000 Value Index, a value-weighted index that roughly omits the largest 1,000 U.S. equities and includes the mid-caps and larger small caps that make up the next 2,000 largest firms. The Russell 2000 Value Index represents those firms within the Russell 2000 Index that exhibit the equity style of value (having lower price-to-earnings ratios and lower expected growth rates). The high observed correlation is consistent with the view of U.S. REITs as having capitalizations ranging from the larger end of the small caps to the smaller end of the mid-caps. Further, REITs are usually viewed as having the style of value rather than growth, owing to their relatively large earnings and small expected growth rates. The high dividend payout ratios of REITs limit the ability of internally generated capital to fund growth rates. Thus, the REIT Index behaves in tandem not simply with the overall stock market but with the sector of the market that is most fundamentally related in terms of size and growth. The NCREIF Property Index indicates slightly higher correlations with value stocks than with growth stocks but does not indicate strong tendencies to correlate with small stocks. A possible interpretation of the correlations is that the REIT Index appears to reflect more economically reasonable behavior than does the NPI.

The unsmoothed version of the REIT Index behaves much like the smoothed version in terms of correlations with Russell indices, and the smoothed and unsmoothed versions have a correlation of 0.99. This is consistent with the low autocorrelation of the REIT Index (0.15), which is attributable to its prices being determined in competitive markets. The unsmoothed version of the NPI, however, has a correlation of only 0.59 with the original NPI, indicating its high autocorrelation (0.85) and the resulting major effect of unsmoothing.

The unsmoothed version of the NPI has a substantially higher correlation with the REIT Index and with the equity indices, indicating that the unsmoothing process removes lags that prevent a more timely recognition of changes in the values of the underlying assets. However, the REIT Index still appears to maintain a substantial edge in the reasonableness of its correlations as being indicative of the accuracy and speed with which it reflects true value changes in its underlying real estate assets.

The REIT correlations in Exhibit 17.11 appear to be intuitively reasonable and consistent. Unfortunately, the high correlations indicate that equity REITs do not offer substantial diversification benefits. The NCREIF correlations in Exhibit 17.11 would appear to indicate substantial diversification benefits if the index were tradable. However, given that the underlying appraised values are not tradable, the results may simply indicate that the appraised values are too smoothed and too lagged to be trusted.

In an analysis of earlier data, Clayton and MacKinnon (2001) find similar results: "The results show that, over the entire 1978–1998 sample period, REIT returns exhibited the greatest sensitivity to bonds and stocks (both small and large cap). After accounting for these public market factors, there was no role for unsecuritized real estate in explaining REIT returns."

The question as to whether real estate serves as an effective diversifier is complex. The answer depends on which, if either, of the real estate indices accurately reflects underlying real estate values: appraisal-based indices or market-based indices. The belief that U.S. commercial real estate equity serves as a powerful diversifier to a portfolio of U.S. equities must be based on two assumptions: (1) that equity REIT prices do not accurately reflect the value of their underlying real estate, and (2) that commercial real estate property values have a low correlation with U.S. equities.

17.5 CONCLUSION

This chapter has reviewed the most important characteristics of unlisted real estate funds (open-end and closed-end funds) and listed real estate products (REITs and ETFs based on real estate indices).

The chapter also offered an empirical analysis of the differences between market-based and appraisal-based real estate returns, and showed that these two valuation approaches offer considerably different indications of the volatilities of real estate returns, thus affecting the accuracy of risk estimations. The divergence between the returns of appraisal-based and market-based real estate indices can be partially attributed to the smoothing caused by appraisals. Indeed, an unsmoothed version of the NPI was shown to have a higher correlation with REITs and equities compared to the standard (smoothed) NPI. Finally, explanations for the divergence between appraisal-based and market-based indices were offered in the context of similar circumstances with other investments.

We have already discussed a number of products and themes related to real estate investing in an international context (for example, we just devoted a section to global

REITs). However, we still need to expand on the opportunities and challenges of international real estate investing, and on establishing a global real estate equity investment program. The next chapter is dedicated to these issues.

NOTE

1. All of the fund names and classifications are from Seeking Alpha, http://seekingalpha.com/article/30370-real-estate-reit-etfs.

REFERENCES

Baum, A., and P. Struempell. 2006. "Managing Specific Risk in Property Portfolios." University of Reading.

Cannon, S. E., and R. A. Cole. 2011. "How Accurate Are Commercial Appraisals? Evidence from 25 Years of NCREIF Sales Data." *Journal of Portfolio Management* 35 (5): 68–88.

Clayton, J., and G. MacKinnon. 2001. "The Time-Varying Nature of the Link between REIT, Real Estate and Financial Asset Returns." *Journal of Real Estate Portfolio Management* 7 (1): 43–54.

Edwards, T. M. and D. F. Bernstein. 2008. "REITs Empowered." *Tax Management Real Estate Journal* 24 (11): 1–14.

Fisher, J., D. Gatzlaff, D. Geltner, and D. Haurin. 2003. "Controlling for the Impact of Variable Liquidity in Commercial Real Estate Price Indices." *Real Estate Economics* 31: 269–303.

Humphreys, T. 2008. "Congress Enacts New Laws Affecting REITs." Morrison & Foerster LLP Client Alert, July 31. Available at www.mofo.com/congress-enacts-new-laws-affecting-reits-07-31-2008/.

Husson, T., C. McCann, and C. Taveras. 2012. "A Primer on Non-Traded REITs and Other Alternative Real Estate Investments." Securities Litigation and Consulting Group, Fairfax, VA.

Mühlhofer, T. 2013. "Why Do REIT Returns Poorly Reflect Property Returns? Unrealizable Appreciation Gains Due to Trading Constraints as the Solution to the Short-Term Disparity." *Real Estate Economics* 41 (4): 814–57.

Parker, D. 2011. *Global Real Estate Investment Trusts*. Oxford, UK: Wiley-Blackwell.

CHAPTER 18

International Real Estate Investments

Cross-border real estate investing offers opportunities for enhanced diversification. Investors often begin real estate investing domestically because local opportunities are more familiar and easier to undertake than cross-border (international) investments. As investors gain experience investing in a variety of domestic real estate projects, they often look to other countries for new opportunities. Institutional investors pursuing direct access to real estate assets often begin investing abroad by becoming a limited partner in an international project somewhat similar to the domestic opportunities with which they are already familiar and experienced.

18.1 OVERVIEW OF INTERNATIONAL REAL ESTATE INVESTING

Historically, institutions have dominated investments in commercial real estate throughout most markets and most countries, with the vast majority of holdings held domestically. The primary problems encountered with international investment include lack of local knowledge, agency costs, regulatory restrictions on foreign ownership, higher transaction costs, complex taxation, small-scale markets, political and economic risks, exchange rate risk, and access to local services. The growth in the use of real estate funds for real estate investments has facilitated an increase in cross-border real estate investments. Real estate funds allow investors who lack the resources or skills to invest directly to enjoy the benefits of diversification into cross-border products without the need to know the characteristics and functioning details of foreign property markets.

The growth in international real estate investments, while widening the potential pool of properties available for purchase, does not come without challenges. In particular, real estate funds are exposed to foreign exchange risk. Currency exposure is particularly apparent in the case of international real estate investments, given the relative low volatility of property operating income. This low volatility of operating income often renders exchange rate movements as the main source of short-term risks and returns (positive or negative) from investing abroad, unless currency risk is hedged.

According to Almond and Vrensen (2014), the total stock of commercial real estate in Europe, the Asia-Pacific region, and North America amounted to $33 trillion as of 2012, distributed as follows:

1. Approximately 37% (around $12.4 trillion) of this total is currently *invested* and is distributed as follows: Europe ($4.4 trillion), Asia-Pacific ($4.25 trillion), and North America ($3.75 trillion). In 2011, the total stock of invested property surpassed the record amount of 2008; and in 2012, it was at an all-time high (in nominal U.S. dollars), with Asia-Pacific and Europe gaining an increasing share of the total stock almost continually during the previous decade. In the case of Europe, this was partially caused by the appreciation of the euro with respect to the U.S. dollar during that decade.

   In terms of capital source, the $12.4 trillion of invested capital had been financed by private equity ($4.2 trillion), public equity ($0.9 trillion), public debt ($1.5 trillion), and private debt ($5.8 trillion).
2. Approximately 26% (around $8.2 trillion) is regarded as *investable* and is currently owner occupied (i.e., held by occupiers and attractive to investors).
3. Approximately 36% is *noninvestable*, which means that this category of commercial real estate is owner occupied and is not available to investors due to use or quality.

Return characteristics, taxes, and transaction costs associated with direct foreign real estate investing can vary extensively from one country to another (see Exhibit 18.1 for developed markets and Exhibit 18.2 for emerging markets). For example, Exhibit 18.1 shows that rental income taxes for foreigners across a group

EXHIBIT 18.1 Return Characteristics, Taxes, and Transaction Costs Associated with Direct Foreign Real Estate Investing: Developed Market

Country	Gross Rental Yield	Rental Income Tax	Roundtrip Cost	Capital Gains Tax
Austria	2.18%	33.92%	11.20%	25.00%
Belgium	5.53%	9.22%	21.10%	n.a.
Canada	3.68%	14.87%	7.85%	25.00%
Denmark	4.84%	3.67%	2.23%	24.00%
Finland	3.63%	23.80%	9.01%	32.00%
France	2.89%	10.00%	18.45%	33.30%
Germany	3.34%	15.82%	12.68%	n.a.
Hong Kong	2.82%	12.16%	34.11%	n.a.
Ireland	n.a.	10.05%	9.07%	33.00%
Japan	5.02%	1.70%	13.36%	15.00%
Netherlands	5.45%	15.00%	12.16%	1.20%
New Zealand	6.09%	1.74%	4.52%	n.a.
Norway	n.a.	26.86%	4.69%	27.00%
Spain	3.91%	24.75%	12.75%	21.00%
Sweden	n.a.	22.76%	8.26%	30.00%
Switzerland	3.81%	48.56%	6.21%	n.a.
United Kingdom	3.21%	n.a.	8.03%	28.00%
United States	3.91%	30.00%	10.00%	5.00%

Source: www.globalpropertyguide.com (accessed on April 14, 2015). The assumptions made to obtain these numbers are explained on the web page.

EXHIBIT 18.2 Return Characteristics, Taxes, and Transaction Costs Associated with Direct Foreign Real Estate Investing: Emerging Markets

Country	Gross Rental Yield	Rental Income Tax	Roundtrip Cost	Capital Gains Tax
Argentina	7.30%	21.00%	10.79%	n.a.
Brazil	4.71%	15.00%	11.50%	15.00%
China	2.66%	5.00%	7.10%	20.00%
India	2.22%	8.11%	11.88%	30.00%
Mexico	7.76%	3.36%	7.89%	25.00%
Poland	6.71%	17.44%	6.73%	n.a.
Russia	3.22%	30.00%	23.11%	30.00%
South Africa	3.88%	7.23%	16.63%	40.00%
Turkey	5.09%	14.61%	8.80%	n.a.

Source: www.globalpropertyguide.com (accessed on April 14, 2015). The assumptions made to obtain these numbers are explained on the web page.

of developed countries can be as low as 1.7% (Japan) and as high as 48.56% (Switzerland). Similarly, capital gains taxes can be as low as 1.2% (the Netherlands) and as high as 33.3% (France). We can also observe a similarly high variability in income tax and capital gains tax rates applicable to foreign investors across major emerging markets (Exhibit 18.2). **Roundtrip costs**, which are the total costs of buying and selling a residential property, including legal fees, sales and transfer taxes, registration fees, and real estate agents' costs and fees, also show a very high variance across countries. Overall, the exhibits make it clear that it is of the utmost importance for investors to understand the specific characteristics of each real estate market in which they might consider allocating funds.

18.2 OPPORTUNITIES IN INTERNATIONAL REAL ESTATE INVESTING

Investing in international real estate offers several opportunities. These include the expectation of higher returns, diversification benefits, and potential tax savings.

18.2.1 Expectation of Higher Returns

One of the reasons for going international is the expectation of achieving higher returns abroad than in the home country. These higher expected returns may originate out of the dynamics of structural variables (e.g., economic growth and demographics) or business cycle considerations. Clearly, not all investors can expect to earn a higher return by investing abroad. In other words, some investors may find expected returns from local real estate investments superior to any that can be found abroad. Real estate is no different from other investments in which investment capital flows to the locations that offer the highest expected returns.

The economies of emerging markets grew at nearly twice the rate of those of developed markets during the past decade, albeit with greater volatility. This trend is expected to continue during the next decades. Property demand should follow this trend, which is also correlated with demographics. Whereas the total population of most developed markets has already peaked (for example, in most of western Europe) and in a few of these countries is even declining, the population in most emerging markets is expected to continue to grow during the next few decades (except in a few cases, such as many eastern European countries and China, where population is expected to peak about a decade from now). On the other hand, migration to a few developed countries (for example, the United States and Australia) has helped the populations of these countries grow, albeit modestly. Economic and population growth in emerging markets has been accompanied by fast urbanization, with the resulting impetus in the demand for new and improved office space, retail space, and housing.

Demand for office space is influenced by the number of people of working age, because they will occupy the offices, and by the level of employment. Demand for retail space is partially affected by total population and its particular composition, because consumption is not the same for all age groups. Furthermore, the purchasing power of the population is a crucial factor affecting demand for retail space. Housing demand mostly depends on household formation, which is in turn affected by population size and average household size.

There may be good investment opportunities outside the home country, especially if the home country has a well-developed and mature property market. These enhanced returns are prone to be expected in the riskier markets. Return opportunities abroad can also be motivated by cyclical factors. One example of this is when the domestic market may be thought to have peaked for the cycle, inducing domestic investors to commit funds to real estate in countries where the cycle is more promising. This assumes that timing is possible, and that the expected return of the strategy will be enough to offset the corresponding transaction costs, which can be quite high, as will be explained later.

According to Geltner et al. (2014), property markets in many countries, particularly in developed markets, are becoming replacement markets in which (1) there will be increasing competition between regions, (2) good quality will drive out bad quality, and (3) the supply will have to bend to mitigate the effect of falling prices. Therefore, investors from demographically mature markets (mostly developed countries) should make strategic allocations to international real estate, especially to demographically immature markets (mostly emerging markets).

18.2.2 Diversification Benefits

Although not every investor can expect to earn a higher rate of return from international real estate investments, investors are likely to benefit from the diversification benefits offered by such investments. Domestic real estate markets are influenced by national and international economic variables. Countries have different economic structures, and this provides a local drive to real estate markets. For example, while demand for real estate in each country is affected by local employment, consumer spending, and local demographics, supply is partly driven by local factors, such as capital availability and its cost. Furthermore, cash flows from real estate are affected by local market institutions, such as rental contracts and zoning rules. For example,

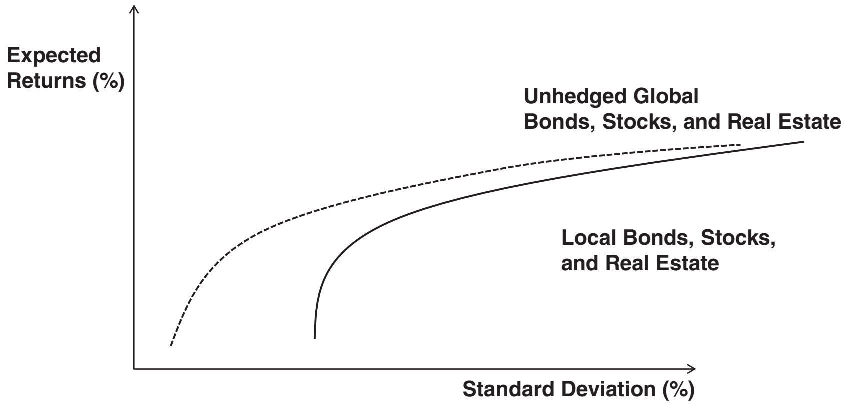


EXHIBIT 18.3 Global (Unhedged) versus Local Efficient Frontiers

Geltner et al. (2014) report large differences in rental contracts (in terms of lease length, right to renew, and so on) around the world.

To the extent that these national economic variables and business cycles are nonsynchronous with the economic factors and business cycles in other countries, one would expect relatively low correlations across international real estate returns. Therefore, a diversified investment in international real estate markets should have a relatively large potential for benefiting investors by offering access to different cash flow streams and risk exposures.

In theory, an efficient frontier that includes global bonds, stocks, and real estate dominates a frontier that includes only local or domestic bonds, stocks, and real estate because it offers a larger investment set of assets to choose from, which are not perfectly correlated. Graphically, this set corresponds to the efficient frontiers depicted in Exhibit 18.3, where for each possible level of expected return the investor can anticipate a lower level of total risk when combining local and international investments rather than investing only domestically.

As we will discuss later in this chapter, one of the greatest challenges to international investments in real estate is currency risk. However, in many cases, this risk can be hedged. Exhibit 18.4 takes Exhibit 18.3 and adds another efficient frontier of global assets, including international real estate, this time with currency risk being hedged. Whereas exchange rates may display long-term trends due to differences in inflation rates in the two countries, much of the short-run variation in exchange rates represents noise around a trend. Therefore, in most cases, for each possible level of expected return, investors can expect a lower level of risk when they go international and hedge the currency risk of their investments, rather than investing internationally without hedging the exchange rate risk. In practice, however, the hedging of currency risk can be quite expensive, as we will discuss later. Also, currency hedging could lead to lower rates of return if foreign real estate, equity, and bond investments take place in countries with strong currencies. In short, although international investments, including real estate, may not always shift the efficient frontier up, they should shift the frontier to the left. In other words, potential diversification benefits should be the major drivers of global real estate investments. Institutional investors are well aware of this fact, as more than 65% of them cite diversification benefits

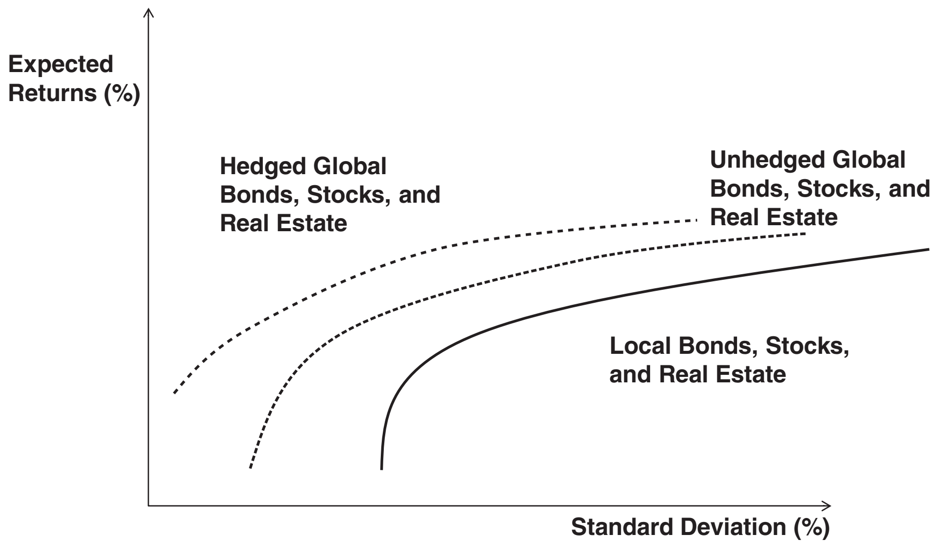


EXHIBIT 18.4 Global (Hedged and Unhedged) versus Local Efficient Frontiers

as the major reason behind international real estate investment, compared to 16% citing higher returns as the main reason (see Preqin 2015).

According to Geltner et al. (2014), the empirical evidence is mostly supportive regarding the benefits of international diversification in real estate. Results depend mostly on the correlations between international real estate returns. Unfortunately, these correlations are not stable, not even for short horizons. Globalization, the emergence of economic blocs, and increasing international capital market integration have led to an upward shift in correlations, thus diminishing the benefits of international diversification (we will discuss more on this at the end of the chapter). However, the problem of structurally increasing international correlations is more serious for stocks and bonds than it is for real estate assets.

Diversification by institutions should be driven by reason and evidence in the pursuit of an optimal portfolio. A starting point for optimal diversification is to begin with a target portfolio equivalent to the market portfolio prescribed by the capital asset pricing model (CAPM), and then to adjust from market weights to optimal weights that include the objectives and constraints of the investor.

The general prescription of the CAPM—that each asset class should be represented in proportion to its total market value—is not valid for either domestic or international real estate investments. In other words, the market weight is generally not an optimal weight for every investor in terms of allocation to real estate. The reason for this is that real estate offers unique advantages (e.g., preferential income tax treatment), which we discuss in the next section, and disadvantages (e.g., illiquidity and lumpiness), which should drive the portfolio allocation decision on an investor-by-investor basis. Also, it is unclear what constitutes the total market value of global real estate. Should it include residential and commercial properties occupied by their owners? Because real estate is lumpy and has characteristics of both an investment and a consumption good, it is unclear even from a theoretical point of view how real estate should be handled in the context of Markowitz's portfolio construction and efficient frontier.

18.2.3 Importance of Income Taxation Analysis

An essential element of real estate portfolio allocation, even for tax-free institutions, is income taxation. The essential idea is this: Investors in high income tax brackets should concentrate on (i.e., overweight) real estate investments that offer substantial income tax advantages, and investors in low or zero income tax brackets should generally avoid (i.e., underweight) investments that offer substantial income tax advantages. This is similar to the idea that tax-exempt investors should not invest in tax-exempt bonds (e.g., municipal bonds).

Pension funds are usually tax exempt in their own domicile but may suffer withholding taxes when investing abroad. There may also be other types of taxes for foreign investors in real estate. For example, there are a few cases of countries that charge an annual flat tax (in percentage) on the value of properties owned by foreigners. There may also be other taxes, such as property taxes and sales taxes. The complexity of the problem makes the use of expert tax advice of the utmost importance. In the next sections, we discuss depreciation tax shields (which are one of the main tax advantages of investing in real estate, and the reason for including income taxation as an opportunity, even though it is also a challenge to international real estate investing); deferral of taxation of gains; and the combined tax effects of depreciation, deferral, and leverage.

18.2.4 Depreciation Tax Shields

The primary potential tax advantage of real estate around the world is the deductibility of depreciation on buildings for income tax purposes. Thus, a property generating net cash revenues from rent and leases of €5,000,000 per year will generate taxable income of only €4,500,000 per year if the taxpayer is able to claim €500,000 per year in depreciation expense.

The tax advantages relating to the deductibility of depreciation for income tax purposes were detailed in the CAIA Level I core curriculum, using the after-tax internal rate of return (IRR). The effect of depreciation was demonstrated using after-tax IRRs and was shown to cause effective income tax rates to differ from stated income tax rates. Specifically, effective income tax rates were shown to depend on the extent to which the depreciation allowed for tax purposes exceeded, equaled, or was lower than the actual decline in the value of the real estate. If the depreciation allowed for tax purposes exceeds the actual economic decline in the value of the real estate (which it usually does), then effective income tax rates can be substantially lower than stated income tax rates.

This analysis of the tax advantages of depreciation focuses on the present value of the depreciation tax shield. The depreciation tax shield is the prospective stream of reduced income taxation that a particular investor will experience as a result of being able to deduct depreciation:

$$\text{Depreciation Tax Shield}_t = \text{Depreciation}_t \times \text{Tax}_t \qquad (18.1)$$

where Depreciation_t is the amount of depreciation that the investor can deduct in year t, and Tax_t is the investor's marginal tax rate in year t. Thus, Depreciation tax shield_t is the amount by which the investor's income taxes in year t are reduced as a result of being able to deduct depreciation at the tax rate, Tax_t.

The present value (PV) of the depreciation tax shield is computed by discounting the stream of cash flows at an appropriate interest rate. Assuming that tax rates are constant and that the investor will be able to enjoy the tax shield with certainty, the cash flows can be discounted at the riskless rate, R_f, as illustrated in Equation 18.2:

$$\text{PV Depreciation Tax Shield} = \sum_{t=1}^{T} \frac{\text{Depreciation}_t \times \text{Tax}_t}{(1+R_f)^t} \tag{18.2}$$

For example, a building offering a €10,000,000 depreciable base to be evenly depreciated over 20 years (using straight-line depreciation) will permit deduction of an annual depreciation of €500,000 per year. At a marginal tax rate of 40%, the annual depreciation tax shield is €200,000 per year. At a discount rate of 5%, the present value of the 20-year tax shield is €2.49 million.

However, the depreciation taken against the building lowers the book value of the building over time. When the real estate is ultimately sold, the taxpayer will typically owe taxes on the profit, including the recaptured depreciation. Thus, the taxable profit on the date of the real estate's sale will be €1 higher for each €1 of accumulated depreciation, and depreciation does not typically change the total nominal amount of taxes paid if tax rates are steady.

Depreciation does, however, reduce the present value of taxes. Assuming that the building in the previous example is sold after 20 years for a value above its original cost, the taxpayer will owe taxes on the recaptured depreciation of €10,000,000. Continuing with a marginal tax rate of 40% and a discount rate of 5%, this tax liability of €4,000,000 in 20 years has a present value of €10,000,000 × 0.40/$(1.05)^{20}$, or €1.51 million. Thus, the net gain to the taxpayer from being able to deduct depreciation is €2.49 – €1.51 = €0.98 million.

If the asset being depreciated for tax purposes is holding steady or appreciating in actual market value, then the ability to depreciate the building may be viewed as an interest-free loan from the government to the taxpayer (i.e., the reduced taxes may be viewed as a loan against future taxes). In other words, relative to not having the real estate and paying full income taxes on all revenues at the stated tax rates, the real estate investor may be viewed as using depreciation as a shield to postpone taxes and increase the discounted value of all future cash flows.

APPLICATION 18.2.4

Consider a real estate investment held by a taxpayer with substantial income, who is in a 25% tax bracket on all income. A property offers $1 million per year in depreciation over the next eight years. However, the depreciation will be recaptured for tax purposes when the property is sold at the end of eight years. What is the present value of the combined tax effects of depreciating the property and recapturing the depreciation using an interest rate of 8.75%?

The benefit is an eight-year $250,000 annuity found by multiplying the annual depreciation amount by the tax rate. The cost is the need to pay taxes at the end of eight years on $8 million of recaptured depreciation (i.e., $2 million). Assuming for computational simplicity that the eighth and final year of depreciation is both deducted and recaptured in the eighth year, the benefits and costs can be computed in a single step with a payment of $250,000, a future value of –$2 million, $N = 8$, and interest rate = 8.75, for a net benefit of $374,314, which may be viewed as the value of an interest-free loan generated by the tax shield.

Depreciation tax shields are offered not only in the United States but also in many other countries around the world. For example, based on a study by Pricewaterhouse-Coopers International, Lee and Swenson (2012) analyze the countries belonging to the European Union and find that all of them offer some type of depreciation tax shield (via either expensing or rapid depreciation).

18.2.5 Deferral of Taxation of Gains

In many countries, real estate can offer taxable investors the advantage of deferred taxation of investment gains until the assets are liquidated. As in the case of common stocks, gains are typically taxed on the sale of the investment rather than being marked to market. Some investments, such as a series of short-term bond investments that are rolled over annually, cause annual taxation of gains.

Annual taxation of gains causes after-tax growth is equal to the product of the pretax rate of gain (r) and 1 minus the marginal tax rate (1 – tax):

$$\text{After-Tax Return without Tax Deferral} = r \times (1 - \text{Tax Rate}) \qquad (18.3)$$

For example, an investment growing at 10% per year experiences an after-tax growth rate of 6% when applying a tax rate of 40%. Thus, a euro placed in a 20-year investment using the same rates would grow to €1 × $(1.06)^{20}$ = €3.2071.

Deferral of taxation of gains is included by compounding an investment forward at its pre-tax rate and subtracting taxes upon liquidation. For example, €1 growing for 20 years at 10%, with gains taxed at 40% at the end of 20 years, would grow as shown:

$$\begin{aligned}\text{After-Tax Future Value of €1} &= €1 \times (1.10^{20} - 1)(1 - 0.40) + €1\\ &= €3.4365 + €1\\ &= €4.4365\end{aligned}$$

Using these numbers, tax deferral allows accumulation of 55.7% more profit: (€3.4365 – €2.2071)/€2.2071. The general formula for the after-tax rate using an annual taxation of gains that is equivalent to earning r for T years with tax deferral is shown in Equation 18.4:

$$\text{After-Tax Return with Tax Deferral} = [1 + [(1+r)^T - 1] \times (1 - \text{Tax Rate})]^{\frac{1}{T}} - 1 \tag{18.4}$$

Note that the annual after-tax return for the preceding example can be found as the rate that equates the after-tax future value (€4.4365) to the initial investment of €1, as indicated in the following equation:

$$\text{After-Tax Rate} = \left(\frac{€4.4365}{1}\right)^{\frac{1}{20}} - 1 = 7.73\%$$

Tax deferral of gains has the effect of increasing the after-tax gains from 6% to 7.73%. Alternatively, note that the pre-tax rate equivalent to 7.73% after tax is found by dividing by (1 – tax), which generates 12.88%. Thus, tax deferral of gains relative to annual taxation of gains is equivalent to being able to earn 12.88% pre-tax rather than 10%.

APPLICATION 18.2.5

An investor is comparing the after-tax rates of return on an investment offering a pre-tax return of 5.2% per year. Inside one investment wrapper, the return is fully taxed each year. Another wrapper defers taxation until the funds are withdrawn. Assuming an investor with a 35% tax rate and a 10-year investment horizon, compute the after-tax returns of the investment using both wrappers.

The wrapper with annual taxation is computed using Equation 18.3 as $5.2\% \times (1 - 0.35) = 3.38\%$. The wrapper with deferred taxation is computed using Equation 18.4 (and a future value of 1.66019) as $[1 + (1.66019 - 1) \times (1 - 0.35)]^{\frac{1}{10}} - 1 = 3.635\%$.

18.2.6 Depreciation, Deferral, and Leverage Combined

As previously indicated, the ability to deduct depreciation as an expense for income tax purposes is, in most cases, tantamount to deferral of income taxes (rather than a reduction in total taxable income). Deferral of income taxes is also available to real estate because capital gains are generally not taxed until realized through a transaction, as discussed in the previous section. Finally, leveraged real estate can offer taxable investors the advantage of deducting interest payments on the debt financing of real estate for income tax purposes in the periods when an interest payment is made.

Taken together, the income tax benefits of leveraged real estate can be especially valuable to investors in very high tax brackets. Returning to the example on depreciation in section 18.2.4, the potential value of the depreciation tax deduction to a taxable investor, almost €1 million in the example, is especially substantial when leverage is applied. If the building in the example (perhaps worth €12,000,000, including land) is financed with 6:1 leverage, the entire €0.98 million depreciation tax shield can be generated with only €2 million of equity investment.

18.2.7 Leverage

Leverage is an integral part of most real estate investing around the world. There are three main reasons for this. First, fund managers will try to use leverage to increase assets under management to improve returns on equity and potentially earn higher performance fees. Second, and as we will discuss later, leverage in foreign currency hedges the debt portion of the investments made. Third, leverage generally offers tax-deductible financing costs for income tax purposes and can be especially valuable to a highly taxed investor trying to enjoy the tax benefits of real estate, such as depreciation.

On the other hand, in analyzing the returns of an equity real estate investment, an investor should be aware of the effects of leverage on volatility. Leverage is a key factor in determining the volatility of an equity position.

Over a short period of time (i.e., ignoring interest expense), and for small changes in value, the returns of a leveraged international equity position, R_{lev}, can be approximated as the product of the return of the underlying assets, R_{assets} (which reflects both the returns of the position calculated in domestic currency and the returns of the exchange rate, which can be hedged, as we will discuss later), and the leverage or gearing factor L (expressed as the ratio of assets to equity), as shown in Equation 18.5:

$$R_{lev} = R_{assets} \times L \tag{18.5}$$

The volatility of the leveraged returns, σ_{lev}, can be derived from the volatility of the asset returns, σ_{assets}, and Equation 18.5, as shown in Equation 18.6:

$$\sigma_{lev} = L \times \sigma_{assets} \tag{18.6}$$

Equations 18.5 and 18.6 ignore the financing costs of leverage for simplicity, since over very short periods of time the costs would be negligible. For example, a rapid drop of 2% in the value of the assets of a position levered 3:1 would cause a 6% drop in the value of the equity. Note, however, that for large changes, the relationship becomes more approximate. Given limited liability, a drop in assets of 40% cannot cause the equity to drop by 120%. Equation 18.6 can be used to approximate that if the volatility of the returns of the assets is 20%, then the volatility of the returns of equity leveraged 3:1 would be 60%. Investors in leveraged real estate should be aware of the potential benefits and costs of leverage, including the impact of changes in asset values over time on leverage.

APPLICATION 18.2.7

An investment with a leverage factor of 3.14 (i.e., the ratio of assets to equity) has a variance of returns of the underlying assets equal to 0.04. What is the variance of the levered equity?

Use Equation 18.6 and note that the asset variance needs to be converted to a standard deviation (volatility) by taking the square root: $L \times \sigma_{assets} = 3.14 \times (0.04^{0.5}) = 0.628$. The resulting standard deviation is converted to a variance of the levered equity by squaring to generate 39.4%.

18.3 CHALLENGES TO INTERNATIONAL REAL ESTATE INVESTING

The growth in international real estate investments, while widening the potential pool of properties available for purchase, does not come without challenges. Here, we examine the following potential obstacles: agency relationships, information asymmetries, liquidity and transaction costs, political and economic risks, exchange rate risk, and legal risks.

18.3.1 Three Reasons Why Agency Relationships Are Important

Institutional real estate ownership typically involves agency relationships, which are important drivers of performance. Agents can fulfill two important functions: serving as managers of the operations of each property, and serving as decision makers in purchasing and selling properties.

The size and nature of a portfolio manager's allocations should be based on the investor's abilities to select, monitor, and manage agency relationships, because the success of the real estate investment is substantially driven by the performance of the managers.

There are three reasons that agency relationships are especially important in the selection and management of real estate relative to the management of other investments, such as passive public equity funds:

1. The market for agents such as real estate managers is not highly efficient. In an efficient market, buyers get what they pay for. For example, in a truly efficient market for athletes, a team owner who hires the most expensive athletes will tend to have the greatest athletes. However, in an inefficient market, buyers cannot be confident that they will earn returns that are on par with the cost and risk of the investment. To the extent that the market for real estate managers is inefficient, an investor's ability to select superior managers with appropriate compensation levels and schemes can justify higher allocations to real estate.
2. In the case of direct property ownership through partnerships with a small number of investors, the investor's relationship with real estate managers may be important, as the real estate investor cannot rely on other investors to monitor

and control the managers. In the case of public equity investing, an investor can remain passive, becoming involved in issues only in limited cases of shareholder activism. Most real estate ownership requires greater contact with real estate managers over the life of the project. Therefore, an investor's ability to have effective ongoing relationships with managers can justify higher allocations to real estate.

3. In a perfectly efficient market, managers cannot consistently generate abnormal profits. Real estate properties often trade in relatively inefficient markets wherein superior managers can consistently generate superior performance. Therefore, investors who can select and monitor managers who are successful in assembling portfolios of real estate properties can justify higher allocations to real estate.

These three aspects of private (i.e., unlisted) real estate emphasize the importance of superior real estate management. Investors must be able to identify capable managers and maintain successful relationships with those managers. The conclusion is simple: Investors with the contacts, access, and capabilities necessary for successful property acquisition and management should consider overweighting private real estate. However, investors unable or unwilling to acquire and maintain superior management capabilities should consider underweighting private real estate.

In general, real estate markets around the world are less efficient than markets for stocks and bonds. This feature creates challenges for investors who want to assess real estate prices and investment prospects. Furthermore, in the case of international real estate investments, the inherent geographical distance between investors and managers will most likely hinder the investor's abilities to select, monitor, and manage agency relationships. One should expect foreign investors to be poorly informed relative to most domestic investors, and this is likely to be reflected in inferior performance. According to Geltner et al. (2014, 626):

> *The information costs (to foreign investors) can come in two varieties, both leading to underperformance. The first is that investors do not have the necessary information and therefore make mistakes. They buy lemons; they pay too much when they buy and get too little when they sell. Alternatively, they could try to solve the problem by buying information, for example from local brokers, or by establishing local offices and employing local people. However, that would simply translate the cost of the information disadvantage into the payment of fees and salaries, likewise eroding the return. In any event, these information costs imply that diversification, dubbed as the only free lunch in financial markets, is no longer free.*

According to results presented in Eichholtz, Schweitzer, and Koedijk (2001) and Eichholtz, Gugler, and Kok (2011), the property market is not efficient, and domestic real estate investors tend to outperform international real estate investors, everything else being the same, when they invest in the same market and product.

Note that investments in large publicly traded real estate markets or other large publicly traded markets do not typically require the ability to select superior managers and to monitor them. In an informationally efficient market, an investor with poor ability to select and monitor managers can earn average market returns by holding a well-diversified portfolio of publicly traded securities. Thus, investors with little or no expertise in real estate or in optimizing agency relationships, such as most foreign real estate investors, may do best by holding well-diversified portfolios of REITs or other market-traded real estate securities. With high levels of diversification and in relatively efficient markets, investor performance should closely track the performance of the overall market without requiring superior management capabilities. However, investors should bear in mind that management fees charged by international real estate funds tend to be higher than those charged by domestic real estate funds, reflecting the higher costs of running a fund that invests abroad.

18.3.2 Information Asymmetries

Due to the uniqueness of most real estate investments, prices of real estate properties offered for sale are more likely to vary widely from their true economic values compared to homogeneous assets trading in relatively efficient markets. The resulting information asymmetries between buyers and sellers can be quite high and cause information costs. In a more efficient market, such as large markets for public equities, there are numerous well-informed buyers and sellers seeking to buy securities that they identify as being underpriced and to sell securities that they perceive to be overpriced. The large number of such informed investors tends to force market prices toward levels that reflect available information. Therefore, a poorly informed investor can be somewhat confident that transactions in efficient markets will generate normal returns.

However, investors with poor information (and analysis) should be concerned that transactions in relatively inefficient markets may systematically generate inferior returns. The net result is that investors with access to superior information and analysis of private real estate should consider overweighting international private real estate, while those with more limited information and analysis should consider underweighting international private real estate.

There exists evidence that domestic investors earn higher returns than foreign investors when both invest in the real estate market of the former. This "home effect" might be explained by the fact that domestic investors are closer to information, and they know more about the local real estate market and the legal and tax situation. According to Baum and Hartzell (2012), this is one of the reasons why foreign investors usually decide to invest through a joint venture with a local partner.

18.3.3 Liquidity and Transaction Costs

Private real estate is typically highly illiquid. The illiquidity of real estate is driven by uniqueness and transaction costs, among other factors. Real estate transaction costs are relatively high. The costs of acquiring and eventually liquidating properties can include sales commissions (agent fees) and legal fees that range between 2% and 7%, depending on the country and the amount of the sale, and other costs, including transfer taxes, search costs, and financing costs, approaching a total of 10%. Property transfer taxes, which are basically nonexistent in the United States, can be relatively high in some countries. They can range from 1% or 2% in Turkey, China, and Mexico to 5% or even 7% in Australia, France, the Netherlands, and Spain.

The substantial length of the real estate sales process is due to the complexities involved. For instance, sellers typically have to prepare the relevant documentation and hire a brokerage firm to market a property, and potential buyers often have to spend a considerable amount of time searching, performing due diligence, and dealing with financing. Furthermore, negotiations between the two parties can take substantial time.

Large interests in private real estate are often restricted to a fixed size that may be inconvenient to traders, thus limiting the liquidity of the interest (i.e., lumpiness). For example, many direct private equity real estate investments take place with a single institutional investor or a small number of institutional investors serving as limited partners. An investor cannot typically adjust the size of the existing limited partnership to increase or decrease his allocation to the interest. Furthermore, in attempting to liquidate the interest, a seller may have difficulty finding a buyer who is both interested in the underlying property and satisfied with the size of the interest being offered.

Illiquidity is a problem that worsens in the case of international real estate investments. This occurs because the foreign investor will be at a disadvantage in dealing with local brokers, especially if she is following an active trading strategy. For example, selling offices in London when the market is thought to have reached a plateau to move to New York City, where it is considered to be at the bottom, cannot be accomplished immediately (as can be done with stocks and bonds), especially if other market participants share the same expectations about the London office market. Illiquidity is higher during periods of weak real estate demand, when the selling of a property can take up to a year or more. Owners usually set minimum reservation prices below which they are unwilling to sell, thus exacerbating the illiquidity problem.

The illiquidity of real estate exacerbates the political and economic risks to which a property may also be exposed, as investors cannot easily liquidate their real estate assets in an environment in which these risks are perceived to be about to become worse in the near future or have already deteriorated. As a result of their illiquidity, it is argued that total returns generated from real estate investments should reflect a liquidity risk premium. A liquidity risk premium is the higher expected return required by investors to compensate them for bearing the risks, costs, and inconveniences of illiquidity. This liquidity risk premium causes private real estate to offer higher expected returns, all else being equal, than are offered by liquid investments. Investors who do not require highly liquid portfolios should consider overweighting illiquid assets, including private real estate, to potentially capture liquidity risk premiums that may be available.

Finally, it is important to point out that some of the obstacles and risks that have been discussed here are less important today than they were in the past, particularly in the case of countries belonging to trading blocs. However, other risks are inevitably intrinsic to international investment in real estate.

18.3.4 Political and Economic Risks

Political risk can be defined very broadly as the risk that returns from investments could be adversely affected as a result of instability or political changes in a country. Political risk is particularly relevant for real estate investments and even more for foreign investors. This is because real estate is part of the heritage of a country, and local governments are key actors in such areas as zoning, taxes, and tenant protection. Some examples of political risk faced by foreign real estate investors are the establishment of new limits on nondomestic property ownership, land and property expropriations, and excessive taxes on foreign investors (Contreras et al. 2014). The illiquidity of directly held real estate investments makes it difficult to escape an expected rise in political risk.

Economic risk is the likelihood that macroeconomic conditions (e.g., changes in monetary and tax policies) and government regulation in a country will affect an investment. In the case of government regulation, there may be restrictions in place on foreigners owning real estate assets. There may also be barriers to the repatriation of profits, an obstacle that may be particularly harmful for foreign direct real estate investments. Some of these issues are gradually becoming less important, especially within trading blocs. We dedicate the next section to discussing a related risk, exchange rate risk, which is of the utmost importance to international real estate investments.

18.3.5 Exchange Rate Risk

Real estate investors are exposed to foreign exchange risk when they invest abroad. Currency exposure is particularly relevant in the case of international real estate investments. This is because a property's operating income has a relatively low volatility, especially when compared to other asset classes, particularly exchange-traded assets such as stocks. This fact renders exchange rate movements as the main factor affecting real estate risks and returns in the short run (from the point of view of an unhedged foreign investor).

A U.S. investor receiving an expected cash flow of €10 million faces two risks in terms of the cash flow measured in U.S. dollars: the volatility of the cash flow, and the volatility of the exchange rate. Similarly, the investor can be viewed as facing two risks regarding the value of the real estate holdings: the volatility of the asset in the local currency, and the volatility of the exchange rate. For example, to the U.S. investor, a property currently worth €100 million can be viewed as being worth the product shown in Equation 18.7 at the end of the next period:

$$€100 \times (1 + r) \times (1 + fx) \approx €100 \times (1 + r + fx) \tag{18.7}$$

where r is the percentage change in the property as valued in euros and fx is the percentage change in the value of euros expressed in terms of U.S. dollars.

APPLICATION 18.3.5a

A European investor purchases a U.S. property and values that property at the beginning of the year as being worth 75 million euros, with the expectation that the property's value will be constant when measured in U.S. dollars. The investor is confident that the property will generate a 4% cash flow during the year but wishes to know the total return (in euros) under two scenarios: (1) if euros appreciate 2% relative to the U.S. dollar, and (2) if euros depreciate 1% relative to the U.S. dollar. Find the total return measured in euros for each scenario.

Equation 18.7 can be viewed as expressing the percentage total return in the value of the property in the investor's domestic currency as the sum of the total return of the property (including cash flows) in the investor's foreign currency plus any appreciation in the investor's home currency relative to the foreign currency (or minus any depreciation). Under the scenario of a 2% appreciation in the euro relative to the U.S. dollar, the total return in euros would be 2%, found as the sum of the U.S. dollar return (4%) and the −2% strengthening in the euro relative to the dollar. Under the scenario of a 1% depreciation in the euro, the total return in euros would be 5%.

The investor's risk can be effectively viewed as a two-asset portfolio if r and fx, property and currency, are both thought of as assets. Analogous to the case of a two-stock portfolio, this international investment's variance can be decomposed as follows:

$$\sigma_d^2 = \sigma_{fx}^2 + \sigma_r^2 + 2\text{cov}(fx, r) \tag{18.8}$$

where σ_d^2 is the variance of the international real estate investment return, expressed in domestic currency terms; σ_{fx}^2 is the variance of the foreign exchange rate; σ_r^2 is the variance of the foreign real estate asset return in its currency; and $\text{cov}(fx, r)$ is the covariance between the foreign exchange rate and the foreign real estate asset return. Equation 18.8 illustrates the idea that an investment in real estate domiciled in a location with a different currency can be viewed as adding foreign exchange risk, especially in instances in which revenues are dominated by leases fixed in terms of the foreign currency and expenses such as interest expense are also specified in terms of the foreign currency.

APPLICATION 18.3.5b

Return to the European investor in Application 18.3.5a, who purchased a U.S. property and is concerned about the total return of the property when viewed in euros. The investor views the U.S. dollar total return as having a standard deviation of 20% and views the standard deviation of the exchange rate between euros and the U.S. dollar as having a standard deviation of 15%. Find the total variance of the euro-based return of the property under two scenarios regarding the covariance between the foreign exchange rate and the foreign real estate's return: (1) covariance = 0%, and (2) covariance = 1%.

Equation 18.8 expresses the variance of the investor's total return viewed in the investor's home currency as the sum of two variances and twice the covariance. First, note that the variances are 0.04 and 0.0225. For scenario 1, the total variance is the sum of the variances: 0.04 + 0.0225 = 0.0625. Under scenario 2, the total variance adds in twice the covariance for a total of 0.0825.

This analysis of foreign investment as having added risk due to the operation of the real estate in a jurisdiction with a different currency is based on measuring risk in the currency of the investor's home country. However, it is not clear that investor wealth and risk should be measured with respect to a single currency. Some institutional investors may have a global investor base. In addition, investors who minimize currency risk measured in terms of a single currency are likely maximizing their risk exposure to the risk of unanticipated inflation in that same currency. A strong argument can be made that investors should consider having their wealth exposed to a diversified basket of currencies to avoid concentration of inflation risk in a single currency.

Real estate investors should consider currency risk in making cross-border real estate allocations. Derivatives allow most short-term currency risk exposures to be hedged if desired. To the extent that an investor perceives diversification into nondomestic currencies to be a plus, cross-border real estate investment may be especially attractive and worthy of higher allocations. However, long-term hedging of currency exposures may be ineffective and costly. For example, long-term forward contracts may not be available, and future forward rates are as volatile as future spot rates. Also, even when such long-term derivative contracts exist, the transaction costs are likely to be rather high, as the risk is being passed to a counterparty that will have to manage it. It is therefore difficult to justify the use of currency derivatives for investors with a long time horizon, considering that empirical evidence suggests that exchange rates fluctuate randomly. Furthermore, in a portfolio that is invested across many countries and currencies, a large portion of exchange rate risk can be diversified away.

Example 1: Suppose that a real estate investment company based in France has just invested £50 million in property in the United Kingdom. The potential gross income of the first year of operations has been estimated at £4,000,000. However, it is unlikely that the property will be fully occupied all year round, and a 10% vacancy loss rate has been estimated. Given the vacancy rate, the fixed and variable expenses of the property for the first year of operations were estimated at £450,000 and £600,000, respectively. For simplicity, assume that income and expenses occur at the end of the year. The current exchange rate £/€ is 1.30. Calculate the estimated return on the first year to the French company, assuming that the property value is expected to be the same a year from now and that the euro is expected to appreciate 1% with respect to the British pound during the first year. The French company has decided not to hedge the currency risk of this investment. Ignore taxes for simplicity.

Answer: From Level I, we know that the effective gross income from the property in the first year is expected to be £4,000,000 – (£4,000,000 × 0.1) = £3,600,000. The fixed and variable expenses of the property were estimated at £450,000 and

£600,000, respectively, for a total operating expense of £1,050,000 for the first year. Thus, the net operating income (NOI) arising from this property in the first year is estimated to be:

$$\begin{aligned}\text{NOI} &= \text{Effective Gross Income} - \text{Operating Expenses} \\ &= £3{,}600{,}000 - £1{,}050{,}000 = £2{,}550{,}000\end{aligned}$$

Therefore, the expected return on the investment during the first year, expressed in British pounds, is equal to £2,550,000/£50,000,000 = 5.10%. If there are no changes in the exchange rate, the rate of return would be same in euro.

However, the French company needs to calculate the returns in euros, given that its domestic currency is the euro. Considering that the euro is expected to appreciate 1% with respect to the pound, at the end of the year the NOI, expressed in euros, will be £2,550,000 × 1.30 × (1 – 0.01) = €3,315,000. This provides an expected return (in euros) equal to €3,315,000/[£50 million × 1.30] = 5.05%. Notice that we could have obtained this same result by multiplying the return obtained earlier and expressed in pounds by 1 minus the expected appreciation of the euro with respect to the pound: 5.10% × (1 – 0.01) = 5.05%. This is the income return. The total return should include the 1% decline in the value of property in terms of euro.

Example 2: Suppose that a real estate investment company based in the UK has investments in commercial real estate in New York City and San Francisco. If the standard deviation of the returns of the £/$ is estimated to be 16%, the standard deviation of the returns of the portfolio of U.S. properties is 11% (when estimated in U.S. dollars), and the correlation coefficient between the £/€ returns and the U.S. property returns is 0.1, calculate the total risk of this investment from the point of view of the UK real estate investment company, assuming that currency risk is not hedged.

Answer: The total risk of this investment from the point of view of the UK company is the variance of the international real estate investment return, expressed in domestic currency terms (i.e., in £ for a UK-based company) or σ_d^2:

$$\sigma_d^2 = \sigma_{fx}^2 + \sigma_r^2 + 2\text{cov}(fx, r)$$

Given that $\rho_{fx,r} = \frac{\text{cov}(fx,r)}{\sigma_{fx}\sigma_r}$, we can find

$$\text{cov}(fx, r) = \rho_{fx,r}\sigma_{fx}\sigma_r = 0.1 \times 0.16 \times 0.11 = 0.00176.$$

Therefore,

$$\sigma_d^2 = \sigma_{fx}^2 + \sigma_r^2 + 2\text{cov}(fx, r) = 0.16^2 + 0.11^2 + 2 \times 0.00176 = 0.04122.$$

and the standard deviation of the international real estate investment returns is 20.30%, which is the square root of the variance 0.04122.

Finally, another obstacle to international real estate investments related to currency risk is the possible existence of capital controls, or the risk that a certain country might impose capital controls when faced with a balance-of-payments crisis (this is also regarded as an example of political risk by some authors). Capital controls make it more difficult and more expensive (in the case of the imposition of a tax for international transactions) to repatriate earnings arising from property investments in the country imposing the controls. This risk is higher in emerging markets.

18.3.6 Legal Risks

Investors in real estate need to verify that, when they purchase a property, they are obtaining a good title, free of encumbrances and liens. In theory, title insurance in the United States can often provide protection against the legal risks of property acquisition. However, legal risks vary from country to country. For example, Girgis (2007) documents that real estate fraud has become a very real menace in places such as Ontario. Specifically, as a result of identity theft, homeowners' titles have been illegally transferred and mortgages have been registered against those titles, all without the homeowners' knowledge or consent. The management of real estate legal risks is closely related to many aspects of the real estate due diligence process.

18.4 ESTABLISHING A GLOBAL REAL ESTATE EQUITY INVESTMENT PROGRAM

A successful strategy for international investments in real estate will benefit from the advantages (return opportunities and diversification potential) and avoid where possible the inherent risks (mainly currency, economic, and political risks) and obstacles (mainly information and monitoring costs). In this section, we first analyze the role of REITs in shaping a global real estate investment program. Then, we examine the role of international direct real estate investments. Finally, we provide a summary of the different forms of real estate investment available—forms that have been discussed in this and the previous four chapters.

18.4.1 International Real Estate Investment Trusts (REITs)

REIT investing allows uninformed investors to make well-informed real estate investments abroad. For example, uninformed investors can buy REIT shares on many stock exchanges, such as those of Australia, France, Hong Kong, London, and the United States. Public markets are generally regarded as being fairly efficient. Furthermore, transaction costs for REITs are low, and investors in these investment vehicles do not have to worry about close monitoring of a portfolio of international real estate assets. And, as we noted previously, tax-exempt structures for listed property companies are becoming the norm around the world.

REIT shares trade on stock exchanges around the world and are therefore generally liquid investments, especially when compared to direct real estate investments. Moreover, foreign investors in REITs are not likely to face discrimination when entering or exiting stock markets abroad. Foreign investors who directly invest in estate properties face economic and political risks. However, foreign investors in REITs are better positioned to cope with these risks because their investments are more liquid.

Given that evidence has shown that exchange rates follow a random walk, it is questionable to attempt to cover currency risks, particularly for long-term investments, given the costs involved in hedging. On the other hand, currency risks increase investment risks and may increase the risk-return profile of investors to unacceptable levels, especially for institutional property investors. In any case, if a REIT desires to cover the currency risk to which the portfolio of properties of the company is exposed, it can take positions in currency derivatives, although they are not available for all markets and can have high bid-ask spreads, making the hedging costly. Currency risks can also be mitigated when a listed property company is large enough that it can spread its real estate portfolio across a number of countries and currencies. Finally, a natural and generally inexpensive hedge consists in using leverage denominated in the currency of each of the countries in which the properties are located. In this case, when the currency of a particular country depreciates against the currency of the country in which the REIT reports its financial statements (e.g., in U.S. dollars), the cash flows generated by the properties and denominated in local currency will decline in dollar terms, but at least some of the costs (the financing costs) of these properties will also be lower in dollar terms. Another way to reduce the currency risk would be to have the leases denominated in the currency in which the REIT reports its results (e.g., in U.S. dollars), a practice that has been observed in some emerging markets, especially in times of macroeconomic distress and high exchange rate volatility.

How does an investor determine optimal portfolio (country) weights? In theory, this question could be answered using modern portfolio theory. However, the Markowitz framework suffers from many theoretical and practical shortcomings when it is applied to real estate. For example, historical time series of returns are scarce in real estate, and their descriptive statistics correspond to period-specific outcomes that will not necessarily hold for the period for which the weights predicted by the model will be applied. Therefore, it would be incorrect to use the Markowitz approach to determine optimal country weights for the case of real estate investments.

A second approach consists of using international index weights to track the composition of the global real estate market. However, the real estate markets of a number of countries are underweighted in global real estate indices, whereas the real estate markets of other countries are overweighted. For example, according to Geltner et al. (2014), Australia happens to have a highly securitized real estate market and is therefore overrepresented in the global property portfolio. A solution to this problem consists in using index weights with gross domestic product (GDP) adjustments. This procedure yields country allocations that are more stable and that are perhaps closer to the true global property market weights.

As we discussed earlier in the chapter, one of the main shortcomings of investing in international REITs to obtain exposure to international real estate markets is the existence of high correlations between REIT returns across countries. For example, Exhibit 18.5 shows the three-year rolling correlation between U.S. and UK securitized real estate market returns between January 1993 and January 2015, based on returns on the Global Property Research (GPR) General Indices of these two countries. Whereas between 1993 and 2007 the rolling correlation fluctuated most of the time between 0.3 and 0.5, the rolling correlation increased dramatically between 2008 and 2013 (coinciding with the global financial crisis and its aftermath), fluctuating between 0.7 and 0.85, and then decreased to around 0.6 in 2014. This behavior highlights that correlations among asset returns change over time, and that they may increase substantially during times of market distress, precisely when diversification across assets is most needed. It remains to be seen whether the correlation between the securitized real estate market returns of these two countries will eventually return to the 0.3–0.5 range that preceded the global financial crisis.

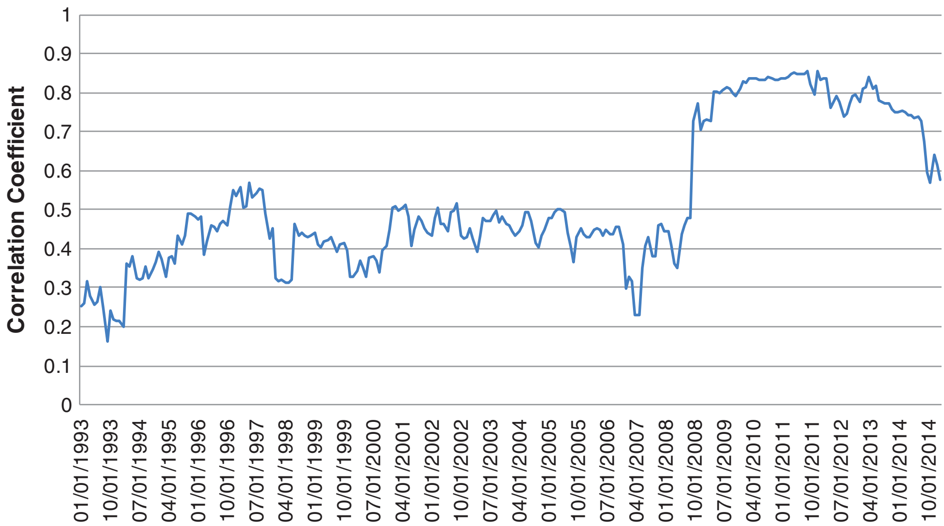


EXHIBIT 18.5 Three-Year Rolling Correlation between U.S. and UK Securitized Public Real Estate Market Returns (Monthly Data)
Source: Global Property Research (www.globalpropertyresearch.com).

Exhibit 18.6 shows the three-year rolling correlation between the securitized real estate market returns of France and Germany, the two largest economies of the euro area, between January 1993 and January 2015. It can be seen that the rolling correlation has been increasing, albeit with great volatility, during this period,

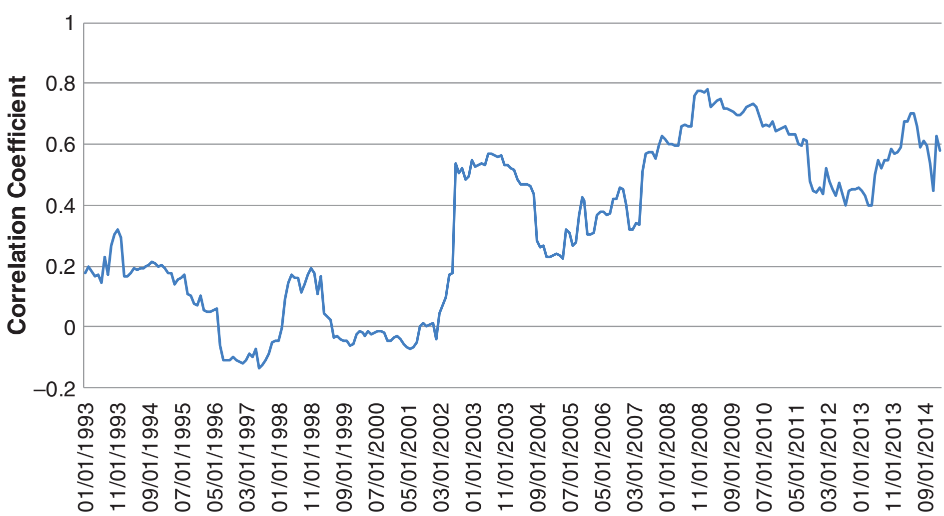


EXHIBIT 18.6 Three-Year Rolling Correlation between French and German Securitized Public Real Estate Market Returns (Monthly Data)
Source: Global Property Research (www.globalpropertyresearch.com).

particularly around the time of the global financial crisis of 2008–09 and its aftermath. The rolling correlation coefficient has increased from a range of between 0 and 0.2 during the first decade to a range between 0.2 and 0.8 between 2002 and 2014. The long-run tendency has been an increase in the correlation, which can be explained by the increasing economic integration between these two countries and the introduction of the euro in 1999, which completely eliminated currency risk between the two countries.

Direct real estate returns, in contrast, exhibit low correlations, as the reported returns of these products are smoothed. Also, if direct property markets are inefficient, investors should focus on their home market, where they have an information advantage. This provides at least some justification for making direct real estate investments, although the challenges of this type of investment must be considered before allocating funds to it.

We conclude by pointing out that property companies should focus their investments on the home market (as long as they have an informational advantage) and thus attempt to outperform the market. Diversification across the portfolio of companies does not add value to a property company; it is the shareholders who should diversify their portfolios by investing across the universe of REITs, stocks, bonds, and alternative investments that may be suitable to them.

End investors—such as pension funds, private persons, and family trusts—should try to find an optimum trade-off between outperformance and diversification. End investors could achieve this by trying to obtain critical mass in home markets and investing in REITs in all other markets, or by concentrating real estate holdings in home markets and investing in stocks and bonds in all other markets.

Example 3: Suppose a U.S. REIT invests internationally. As part of its global real estate investment program, and eager to enhance the expected returns of its investments, the REIT has decided to increase its allocation to emerging markets. The REIT is considering buying an office complex in South Africa. The potential gross income and operating expenses of the first year of operations have been estimated at 80 million rand (R) and 35 million rand, respectively. For simplicity, assume that these amounts are received at the end of the year. A 10% vacancy loss rate has been assumed for this investment. Assume also that the REIT expects to maintain the office complex for five years, rents and operating expenses are estimated to increase by 8% per year, and the vacancy loss rate will remain constant at 10%. Finally, assume that the office complex is projected to be sold in five years, with net sales proceeds estimated at R 280 million. If the REIT is offering R 260 million to acquire the office complex, what is the internal rate of return (IRR) of this investment? Should the REIT invest in this project? Suppose that the required rate of return for this real estate investment is 13% (based on the rand). Ignore taxes for simplicity.

Answer: First, we need to project the net operating income for the next five years. This is shown in Exhibit 18.7, where all income is expressed in rand (R).

We can now calculate the internal rate of return using a financial calculator or Excel (the cash flows are CF0 = –260,000,000; CF1 = 37,000,000; CF2 = 39,960,000; CF3 = 43,156,800; CF4 = 46,609,344; and CF5 = 330,338,092, which includes the proceeds from the sale). The IRR is 17.38%, which is greater than the

EXHIBIT 18.7 Projected Net Operating Cash Flows

	Year 1	Year 2	Year 3	Year 4	Year 5
Potential gross income	R 80,000,000	R 86,400,000	R 93,312,000	R 100,776,960	R 108,839,117
Vacancy loss	R 8,000,000	R 8,640,000	R 9,331,200	R 10,077,696	R 10,883,912
Effective gross income	R 72,000,000	R 77,760,000	R 83,980,800	R 90,699,264	R 97,955,205
Operating expenses	–R 35,000,000	–R 37,800,000	–R 40,824,000	–R 44,089,920	–R 47,617,114
Net operating income	**R 37,000,000**	**R 39,960,000**	**R 43,156,800**	**R 46,609,344**	**R 50,338,092**

13% required rate of return for this real estate investment. Therefore, the REIT should invest in this project.

18.4.2 Summary of Forms of Real Estate Investment

Here we provide a summary of various types of both equity and debt investments that are related to real estate. These forms of investments have been discussed in the Level I book as well as in the previous four chapters.[1]

Equity Investments

- **Direct real estate investments** are investments in physical real estate with little or no pooling or structuring. The investment vehicle is a separate entity and segregated account that provides direct ownership of assets. For international investments, it might be advantageous to enter into a joint venture with local partners. Depending on the style of the investment (i.e., core, value added, or opportunistic), the total return from this investment could be dominated by rental income or price appreciation.
  - ADVANTAGES: Investors can target specific markets or property types. The investment represents the direct ownership of a hard asset. There will be relatively well-defined and stable cash flows from rental properties.
  - DISADVANTAGES: There is exposure to specific property risk because of higher concentration, and there is a need for a property manager and an asset manager. The investment is highly illiquid. Transaction costs are high. The investor must satisfy the highest level of expertise and size. Transaction costs are particularly high for international investments.
- **Pooled investments in direct real estate investments** include investments in a fund that purchases physical property. The investment vehicle is ownership in a pooled vehicle (i.e., a fund), which is not publicly traded (i.e., unlisted). Depending on the style of the investment (i.e., core, value added, or opportunistic), the total return from this investment could be dominated by rental income or price appreciation.
  - ADVANTAGES: They are more diversified than direct investments, and offer more liquidity compared to direct investments. The investment could be smaller in size, and a much lower level of expertise is needed. This is the

most common method by which most large institutional investors invest internationally.
  - Disadvantages: Liquidity is lower than in most REITs. Higher fees are paid to fund managers. Pooled investments are less transparent than direct investments.
- **Listed real estate securities** involve the purchase of shares in publicly traded firms, such as REITs. Historically, the total return on this investment tends to be split between dividend income (about two-thirds) and price appreciation (about one-third).
  - Advantages: They are very easy to diversify domestically and internationally, and offer daily liquidity and pricing and regular transparent reporting. They tend to offer attractive yields.
  - Disadvantages: There is higher correlation with the overall stock market and therefore lower diversification benefits. There is lower total return than direct ownership because they do not provide a liquidity risk premium.

Debt Investments

- **First mortgage claims:** are loans that are backed by the property and have the first claim to the property in case of bankruptcy and liquidation. The investment can take place through direct ownership of the loans or ownership of a pool that invests in such loans. The total return from this investment is based entirely on its interest income.
  - Advantages: The investment has first claim to the properties that back the loan, and provides steady income throughout the life of the loan.
  - Disadvantages: The investments are less liquid than REITs. The value of the investment has greater exposure to interest rates.
- Commercial mortgage-backed securities (CMBS) are securities that are backed by a pool of mortgages. The investment can take place in the form of direct ownership of these securities or investment in a fund that invests in these securities. The total return from this investment is based mostly on its interest income. However, during periods of declining interest rates, price appreciation will be important as well.
  - Advantages: These securities offer various tranches, with each tranche offering specific risk-return profiles, which allow investors to choose securities most suitable for their risk-return preferences. They are more liquid than investments in mortgage loans. A well-diversified pool of mortgages offers greater diversification of risk. These securities offer credit enhancement and therefore have lower exposure to credit risk.
  - Disadvantages: There is less transparency about the pool of mortgages than in investments in mortgage loans; therefore, risk measurement and management (e.g., prepayment risk) might present a challenge.
- **Real estate mezzanine debt** consists of debt instruments that represent a middle position in the capital structure of a real estate property. They appear as subordinated debt or preferred equity positions. The investment vehicle may be in the form of direct ownership of the loans or shares in a pool of capital (i.e., a fund). The total return from this investment is based mostly on its interest income. However, during periods of declining credit spreads, price appreciation will be important as well.

  - ADVANTAGES: These investments have a higher return than CMBS and first mortgages.
  - DISADVANTAGES: They have greater exposure to credit risk and are less liquid than CMBS.
- **Corporate debt of REITs** represents fixed-income securities that are issued by REITs and are similar to debt issued by other corporations. Investors can purchase these directly or through pools of capital (i.e., funds). The total return from this investment is mostly based on its interest income. However, during periods of declining interest rates and credit spreads, price appreciation will be important as well.
  - ADVANTAGES: This investment offers greater liquidity than other types of real estate debt.
  - DISADVANTAGES: These are unsecured and therefore have greater credit risk. They do not represent direct or indirect ownership of real assets. They have greater interest rate risk and tend to be more volatile than other debt instruments.

18.5 CONCLUSION

This chapter has reviewed several distinguishing characteristics of real estate that may be considered by an asset allocator in setting the target portfolio weights for international investments in real estate. The idea is that investors should consider their own tax situation, their tolerance for illiquidity, their ability to manage projects or managers in other countries, and so forth when making allocation decisions for international investments.

Additional issues with international real estate investing should also be considered, such as currency risk exposure and economic and political risk. Furthermore, the challenges of accurately estimating true historical volatilities and correlations, especially in countries in which long histories of returns are not available, substantially increase the risk measurement risk of real estate. **Risk measurement risk** is the economic dispersion caused by inaccuracies in estimating the volatilities and correlations of investments. The challenge of inferring volatilities and correlations for real estate using historical data increases the unreliability of applying state-of-the-art portfolio management techniques, such as mean-variance optimization.

With the complexities and challenges of international real estate investing come opportunities and the potential for enhanced diversification benefits. Matching the returns of traditional investments such as stocks and bonds can be easily accomplished using passive portfolio management approaches, such as indexation. But optimizing the risks and returns of a well-diversified global portfolio of private and public real estate with the liquidity and taxation characteristics appropriate for a particular investment requires expertise and effort. It is especially within the challenging area of real estate that investment professionals can add value.

NOTE

1. See Briddell (2010).

REFERENCES

Almond, N., and H. Vrensen. 2014. *Money into Property: Global 2013*. DTZ Research, www.dtz.lt/uus/failid/File/DTZ+Money+into+Property+2013+Global.pdf.

Baum, A. E., and D. Hartzell. 2012. *Global Property Investment: Strategies, Structures, Decisions*. Chichester, UK: Wiley-Blackwell.

Briddell, E. T. 2010. *A Guide to Global Real Estate Investment Options*. BNY Mellon Asset Management.

Contreras, V., U. Garay, M. A. Santos, and C. Betancourt. 2014. "Expropriation Risk and Housing Prices: Evidence from an Emerging Market." *Journal of Business Research* 67 (5): 935–42.

Eichholtz, P., N. Gugler, and N. Kok. 2011. "Transparency, Integration, and the Costs of International Real Estate Investments." *Journal of Real Estate Finance and Economics* 43 (1): 152–73.

Eichholtz, P., M. Schweitzer, and K. Koedijk. 2001. "Testing International Real Estate Investment Strategies." *Journal of International Money and Finance* 20 (3): 349–66.

Geltner, D., N. Miller, J. Clayton, and P. Eichholtz. 2014. *Commercial Real Estate Analysis and Investments*. Mason, OH: OnCourse Learning.

Girgis, J. 2007. "Mortgage Fraud, the Land Titles Act and Due Diligence: The *Rabi v. Rosu* Decision." *Banking and Finance Law Review* 22:419–34.

Lee, N., and Ch. Swenson. 2012. "Are Multinational Corporate Tax Rules as Important as Tax Rates?" *International Journal of Accounting* 47:155–67.

Preqin. 2015. *Real Estate Spotlight* 9 (6). www.preqin.com/docs/newsletters/re/Preqin-Real-Estate-Spotlight-September-2015.pdf.

CHAPTER 19

Infrastructure as an Investment

This chapter presents an overview of infrastructure as an investment from a variety of perspectives, including various methods of classification, access, regulation, and performance.

19.1 INFRASTRUCTURE ASSETS

Opinions differ as to what constitutes an infrastructure asset of a country or an economy. This section begins with a discussion of how infrastructure can be distinguished from other assets, and then considers how infrastructure assets themselves can be categorized.

19.1.1 Distinguishing Infrastructure

Infrastructure assets are a means for ensuring the delivery of goods and services that promote prosperity and growth and contribute to quality of life, including the social well-being, health, and safety of citizens and the quality of their environments. For example, it may be argued that banking and financial systems are part of infrastructure in view of their being essential to the social and economic fabric of a country.

For the purposes of considering infrastructure as an investment class, it is generally accepted that an asset must meet certain criteria to be included. It must:

- Comprise large, stationary, physical assets with a long and useful economic life.
- Have high capital costs.
- Have low operating costs once construction is complete (e.g., high earnings before interest, taxes, depreciation, and amortization [EBITDA] margins).
- Support the overall functioning of society.

19.1.2 Three Approaches to Classifying the Economic Nature of Infrastructure

Infrastructure assets can be categorized using different approaches. One approach classifies them according to who the payer is:

- The end user pays to use the assets. Examples of these are toll roads, utilities, and communications networks.

- Government and taxpayers pay for the use of assets. In this case, the asset is mostly or partially free for the end users. Examples of these are certain schools and hospitals, administrative buildings, and parks.

Another approach is based on whether the price or cost of using the asset is regulated:

- Regulated pricing: In most countries, the pricing for essential goods and services is heavily regulated, and therefore price changes must be approved by public entities.
- Unregulated pricing: There is a growing trend toward unregulated pricing of certain infrastructure investments in mature countries. These are most common in the energy sector.

The final approach considers the role infrastructure investments play in the economy. In this case, they can be classified as follows:

- **Economic infrastructure assets** are assets with economic value that is driven by the revenue they generate. Typically, end users pay for the services provided by these assets. Examples of these are toll roads and bridges, railways, airports, and maritime terminals.
- **Social infrastructure assets** are assets that have end users who are unable to pay for the services or that are used in such a way that it is difficult to determine how many services were used by each person. Examples of these are schools, public roads, prisons, administrative offices, and other government buildings. Where social infrastructure is provided by the private sector, it is almost always under some form of public-private partnership (PPP).

The degree of private investment within infrastructure depends on the commercial viability of the sector or subsector as well as the governmental policy within a given country. Private-sector participation in economic infrastructure relies on a government that is content with essential assets being operated by the private sector, which charges citizens for its services (often within a regulated pricing structure). It also relies on customers' propensity to pay. For example, toll roads have been historically unpopular with many electorates; consequently, new roads are often procured as "shadow toll" PPPs, meaning that motorists do not pay the tolls. In developing countries, users may be unable to pay, resulting in the need for government support. Private investment in social infrastructure depends both on the government's ability to pay the private sector for the service and on the government's political support for PPP.

19.1.3 The Demand for Infrastructure Assets

The longer-term future performance of the global economy will depend to a large extent on the availability of adequate infrastructure to sustain growth and social development. Through 2030, annual infrastructure investment requirements for electricity, road and rail transport, telecommunications, and water are likely to average around 3.5% of world gross domestic product (GDP) (see OECD 2008).

The Organization for Economic Cooperation and Development (OECD) reports that there is a significant gap between the public sources of funding and the infrastructure investments needed to help the global economy reach its growth potential. Two key questions were raised by the OECD study:

1. Where will the new sources of finance come from, and what role will the private sector play?
2. Will the financial, organizational, institutional, and regulatory arrangements (the "business models") currently in place be able to respond adequately to the complex challenges they face, and are they sustainable over the longer term?

Bridging the infrastructure investment gap will demand innovative approaches, both to finding additional financing and to using infrastructure more efficiently and more intelligently through new technologies, demand management strategies, regulatory changes, and improved planning.

Clearly, there will be growing opportunities for the private sector to invest in infrastructure projects in developed and developing countries. This chapter discusses infrastructure from an alternative investment perspective. It outlines the rationale for investing in the infrastructure of developed countries, and describes the global market opportunity segmented by geographic region. It also touches on regulatory and public policy issues and their effects on the investment opportunity set. The risk-return characteristics and developing trends of infrastructure are also presented with a view to the potential role of infrastructure in investment portfolios. Finally, the chapter discusses how investors can access these investment opportunities, and compares infrastructure investments with similar alternative investments, such as real estate and private equity.

Investing in infrastructure involves applying private sources of capital to achieve meaningful public ends by investing in equity, equity-linked, and debt instruments of businesses that design, build, operate, own, or have long-term concessions in physical assets. These physical assets are those that provide essential services to a society. Good infrastructure facilitates the growth of business, promotes trade, and enhances economic welfare by improving access to vital resources. This, in turn, furthers demand for infrastructure.

Infrastructure businesses have traditionally been within the purview of government ownership and public finance. While the need for infrastructure investment remains high and continues to grow, public funding for infrastructure has declined considerably over the past few decades in the absence of a commensurate increase in taxes. As a result, historical underinvestment, accompanied by a diminished role of government in many countries, is leading to increased reliance on the private sector. This is being addressed through privatization and the outsourcing of new projects. Regulatory and structural changes in North America, the UK, western Europe, and Australia are further catalyzing this change. Formal contracting structures through active public-private partnership programs as well as policy shifts toward "maximizing value for money" propositions typify these changes. Public-private partnership is a government service or private business that is funded and operated through a partnership of government and private sectors. PPP involves a financial and operating relationship between a government entity and a private entity, in which the private entity provides a public service or project and assumes substantial financial, technical, and operational risk for the project. The growth of PPP has led to a convergence of public- and private-sector financing and operational participation in infrastructure businesses.

The large and dramatically expanding financing needs for infrastructure investments in mature markets, let alone those in developing countries, is creating a substantial investment opportunity. This demand is drawing large and growing global capital inflows. Institutions that are not currently invested in infrastructure are undertaking efforts to quickly assess this opportunity and put into place policies, dedicated investment teams, and allocation guidelines. Institutional investors in this area include dedicated infrastructure funds, insurance companies, pension funds, sovereign wealth funds, private equity funds, divisions of large construction companies, and banks. Financial investors are also forming alliances with multinational infrastructure operating companies to invest in infrastructure opportunities.

Infrastructure investments have different risk-return profiles from traditional investments. These differences can be a source of alpha as well as a valuable diversification tool for investors. Typically, demand for infrastructure services is largely inelastic and is relatively unaffected by the business cycle, resulting in a relatively low correlation between returns on infrastructure and market returns.

The United States, Canada, the UK, western Europe, and Australia are at different stages of maturity when it comes to deploying private financing vehicles for meeting public ends. However, they all share one thing in common: namely, that the market opportunity over the next decade is significant.

19.1.4 The Supply of Infrastructure Assets

With few exceptions, underinvestment in infrastructure has been a global secular trend. In some instances, the true economic cost is not passed on to consumers. In these situations, underpricing tends to be subsidized by the government, which results in pressure on government finances. The cumulative adverse impact is underinvestment, deferred maintenance, and inadequate infrastructure replacement. This in turn slows development, erodes productivity, and in some cases results in shifts of populations and industries to geographic areas with more efficient infrastructure.

Governments have historically used bond financings, custom lease structures, special tax districts, tax incentives and credits, and usage fees to facilitate funding of infrastructure projects. They are now increasingly turning to private capital to supplement or replace public financing. This trend began in Australia and the UK and carried over to Canada and continental Europe. Governments can access private capital through outright privatization, PPP, or project finance. This is now becoming increasingly important in the United States. Significant policy work remains to be done in most countries to expand private-sector involvement in infrastructure, although many governments are increasingly focused on advancing such policies. For example, in the United States, a number of drivers are catalyzing increased private-sector involvement. These include (1) funding shortfalls at each level of government caused by limits on tax increases and available debt, (2) divestitures of existing infrastructure assets to raise capital for new investments, (3) initiatives to obtain private-sector management and technical expertise to improve service efficiency, (4) enactments of favorable PPP legislation by federal and state governments, and (5) availability of debt and equity financing from private investment sources.

PPP should be distinguished from project finance. As mentioned previously, PPP is a government service or private business that is funded and operated through a partnership of government and private sectors. On the other hand, **project finance** is the long-term financing of projects, such as infrastructure, wherein the loans are supported by the cash flows from the project. In a typical project financing structure, there are a number of investors or sponsors that provide loans to the operation. The loans are typically non-recourse loans that are secured by the project assets, while the financing costs are covered entirely by project cash flows. Financing is typically secured by all of the project assets, including the revenue-producing contracts. Project lenders are given a lien on all of these assets and are able to assume control of a project if the project company has difficulty complying with the loan terms.

Estimations and methodologies to determine infrastructure funding needs differ due to varying definitions of infrastructure as well as the level of subjectivity involved in assessing the need to renew, maintain, or add to existing assets. The market for private investment in public infrastructure is expected to continue to grow in size, sophistication, and opportunity.

Private investors recognize this trend and have begun to raise dedicated infrastructure funds to address these opportunities. These infrastructure funds bring together capital from a variety of investors and allow for efficient deployment of equity capital.

19.2 STAGE, LOCATION, AND SECTOR OF INFRASTRUCTURE

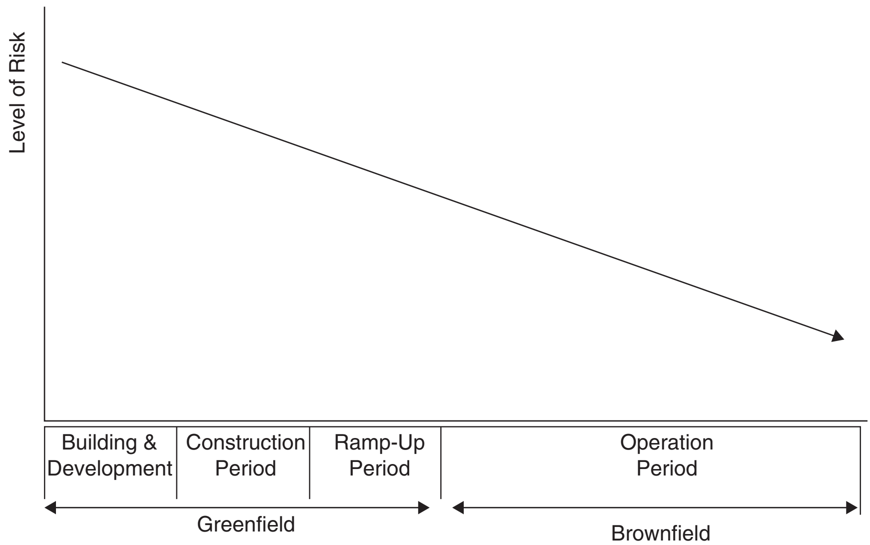
This section describes three important characteristics of infrastructure that largely dictate risk-return characteristics: stage of maturity, geographic location, and sector scope. These three characteristics may be used to broadly classify infrastructure assets for the purpose of analyzing the risk and return of an individual asset or a portfolio.

19.2.1 Stage of Maturity

Investment in assets under construction is riskier than investment in completed assets; investment in newly completed assets with no operating history (including usage) is riskier than investment in mature assets with established operations and usage history. Depending on its phase, an infrastructure investment may be referred to as greenfield or brownfield. The greenfield phase covers the initial stages of infrastructure, from the design to the construction of the project itself and its start-up. By their nature, infrastructure assets tend to involve major construction work, in which risk can be quite high. The brownfield phase takes place when assets are already constructed and, therefore, have a history of operations that provides good visibility into revenue, usage rates, and operating costs (see Exhibit 19.1 for a risk profile of these phases).

As just stated, the greenfield phase of a project contains many risks. The phase can be long and complex and require the coordination of many participants. At this stage, the risks can range from technical to political. Environmental issues are particularly important at this stage, as investors and lenders are becoming increasingly sensitive to these issues. The following is a list of potential risks that are likely to

EXHIBIT 19.1 Risk Profile of Infrastructure Investment Development Stages
Source: World Economic Forum (2014).



arise in the greenfield phase of a project (see Af2i and J.P. Morgan Asset Management 2011):

- Design and technological risks: The design risk may be relatively limited for classic types of projects but can be important if unproven technologies are to be deployed.
- Construction risks: The main risk in this area is that the final cost may significantly exceed the original estimate. This could result from making optimistic assumptions at the beginning or from unexpected changes in the economic environment (e.g., higher than expected inflation or exchange rate changes that increase labor and material costs).
- Economic, legal, and political risks: These risks may be present through the entire life of the project, but they are likely to be most impactful during the greenfield phase of the project, which is when they are most likely to appear.

While the brownfield phase of a project is believed to be less risky, there are certain risks that show up at this stage:

- Revenue risk can arise if the level of use is overestimated. For example, the level of traffic of a toll road may turn out to be much lower than anticipated because drivers are unwilling to pay the extra cost of using the road. Or the output of the project may have been overestimated (e.g., the power plant does not produce the predicted level of energy).
- Operational and maintenance risk is typically low for most infrastructure investments. However, in cases of technological obsolescence, the costs could arise much sooner than anticipated and could come at a much higher level.

EXHIBIT 19.2 Infrastructure Investments: Mature versus Emerging Economies

	Mature Economies	Emerging Economies
Nature of market	Ownership transfer; needs largely related to the maintenance, improvement, and expansion of returns (i.e., refinancing and operational efficiency)	Infrastructure assets yet to be built; fragmented ownership of assets offering consolidation opportunities and multiple exit opportunities
Related risks	Developed legal systems and regulated environments; moderate returns on mature and noncyclical infrastructure	Developing legal and regulatory environments; high perceived risk, requiring higher returns
Operational characteristics	Large-scale projects generally undertaken via competitive tendering	Partnerships with local entities having appropriate industrial and regulatory expertise

Source: Af2i and J.P. Morgan Asset Management (2001).

19.2.2 Geographical Location

Political risk in developing countries is a key consideration given the essential nature of the assets to the local economy, the long life of the assets, and the fact that they cannot be moved, making them vulnerable to expropriation. To mitigate this risk, many investors target only developed countries or developing countries with robust legal frameworks, particularly with respect to property protection and contract enforcement. The long payback period also increases the probability that an investment may be adversely affected by a period of severe economic instability (e.g., government default or hyperinflation). In compensation, many emerging markets are enjoying significant economic growth, which has strong direct benefits for infrastructure businesses. Exhibit 19.2 provides the main characteristics of infrastructure investments in mature versus emerging economies.

19.2.3 Sector Scope

All else being the same, infrastructure businesses whose revenues are not subject to price or volume variations are less risky than those that are subject to variations. Some social infrastructure PPP companies (e.g., for hospital facilities) may be paid for by the government on an "availability basis"; for example, revenue is independent of usage and the service is compensated at a fixed price, with deductions only for poor service. Where there is volume risk, this may be mitigated by the nature of the service (e.g., provision of water to a given area) or by the monopolistic nature of the asset (e.g., being the only airport for a major city). Price risk may be mitigated by regulation that periodically determines price to provide a certain return on capital (e.g., for utilities) or by fundamental demand (e.g., for a toll road for a major corridor). To the extent that price and volume risks are not significantly mitigated, the asset may not be considered to be infrastructure for investment purposes. For example, power generators that sell electricity into a market at spot prices ("merchant" generation, with no long-term power purchase agreements) are considered ineligible by many infrastructure investors.

19.3 TWELVE ATTRIBUTES OF INFRASTRUCTURE AS DEFENSIVE INVESTMENTS

The fundamental drivers of infrastructure performance include changes in macroeconomic factors such as GDP and demographic trends. Growth in these factors increases usage-based revenue and positive cash flows for these assets. Most infrastructure assets feature a set of common attributes that provide for defensive investment characteristics.

19.3.1 Inelastic Demand

Most infrastructure assets and businesses provide essential services that support the functioning of society and the economy, such as power, water, and basic transportation. The indispensable nature of most infrastructure investments results in their demand being relatively inelastic to price changes and economic downturns; their long-term growth is generally proportional to overall economic growth.

19.3.2 Monopolistic Market Positions

More often than not, infrastructure assets and businesses are natural monopolies with high barriers to entry. For example, suppose an airport is the only airport for a particular city. It may be uneconomical to build a competing airport, it may be difficult to get planning approval, and the government may explicitly promise the existing operator not to permit the development of a competing airport in the catchment area.

19.3.3 Regulated Entities

Given the monopolistic nature of such infrastructure assets, governments (or government-sponsored agencies) typically regulate their activities and pricing to preclude undue monopolistic practices and extra-market returns at the expense of the consumer. As regulated entities, they are often required to sell their services at approved tariffs, which are intended to generate sufficient revenues to fund operating costs plus a certain return on capital. Gas, electric, and water utilities as well as transmission assets are examples of these regulated businesses. Their effective management necessitates specialized understanding of the applicable regulatory framework as well as technical and industry expertise to generate attractive risk-adjusted returns. Under the right management, regulated assets may be particularly attractive investments because price regulation mitigates downside risk if costs increase. This is because prices will be allowed to increase to maintain the target return rate. Moreover, returns can still be improved by outperforming the regulator's assumptions for operating costs, capital costs, and cost of capital. On the other hand, because regulated prices tend to be more stable, they tend to be lower than unregulated prices. In addition, while improved services resulting from capital spending can lead to higher prices in unregulated investments, such increases are less likely to occur in regulated industries. Finally, depending on the provisions of the long-term contract, regulated prices may not be allowed to rise fast enough to reflect the increased cost of operations when these costs increase faster than the overall rate of inflation.

19.3.4 Capital-Intensive Setup, Low Operating Costs

While infrastructure assets are capital intensive to set up (e.g., airports, bridges, and tunnels), once established, they generally have relatively low operating costs. This provides for strong operating margins. This attribute, combined with long projected service lives, can support high levels of leverage.

19.3.5 Low Volatility of Operating Cash Flows

In most instances, infrastructure investment revenue streams are relatively stable and predictable, often resulting from either a captive customer base or long-term contracts, regulated pricing schedules, and limited competition or licensing. Utilities are one example. Stable cash flows can also support value-enhancing financial leverage at a more attractive cost of debt capital than similar financings undertaken by more risky assets, such as private equity buyouts.

19.3.6 Resilience to Economic Downturns

Due to their essential role in the economy, infrastructure businesses, once operational, are less likely to suffer from a significant permanent decline in demand, traffic, or patronage than are businesses in other industries. They are expected to better weather downturns, with the possible exception of cases in which inappropriate capital structures have been used to finance their development (e.g., too much leverage) or when demand forecasts have been grossly inaccurate.

Infrastructure investments have varying levels of sensitivity to economic conditions and business cycles. To the degree that returns from certain infrastructure investments are not sensitive to business cycles, they will help reduce the volatility of an investor's overall performance during various stages of a business cycle. Exhibit 19.3 illustrates the general economic sensitivity of various infrastructure investments. Typically, if the cash flow of an investment is contractually fixed, the value of the investment will tend to be less sensitive to changes in economic conditions.

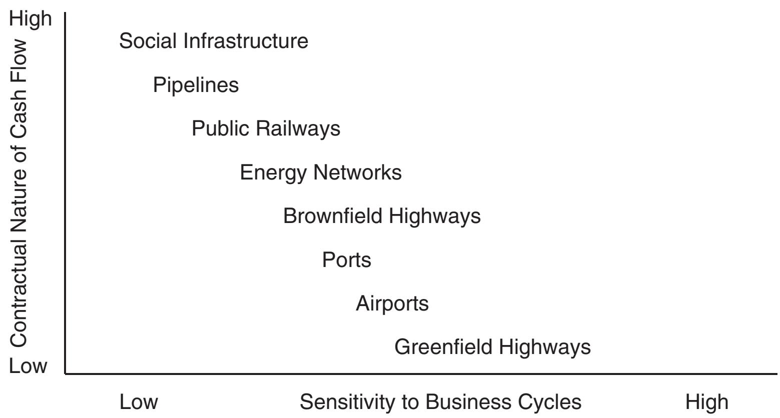
19.3.7 Technology Risk

Some infrastructure assets are less exposed to technology obsolescence risk. For greenfield infrastructure that is reliant on technology (e.g., power generation), it is important to ensure that the technology used is proven or that the risk of technology failure is otherwise mitigated.

19.3.8 Long-Term Horizons

Infrastructure assets have long and useful economic lives (often over 50 years), producing stable revenues with relatively stable cash flows. This long-term predictability

EXHIBIT 19.3 Sensitivity to Business Cycles



may be helpful to investors seeking to match long-term yielding assets against long-term liabilities (e.g., pension plan and insurance company liabilities).

19.3.9 Inflation-Indexed Cash Flows

Infrastructure assets may have contractual or regulatory revenue structures that are adjusted for changes in measures of inflation, such as the Consumer Price Index, thus making for an effective inflation hedge. Their long-term inflation-linked cash flow characteristics are attractive duration hedges for long-term liabilities.

Foreign infrastructure investment projects expose investors to currency fluctuations. If cash flows are contractually fixed, this will make future cash flows more predictable and may reduce the cost of financing these projects. However, the real value of the cash flows will suffer if there is an unexpected increase in inflation. Further, to the degree that the local currency may depreciate in response to higher inflation, foreign investors will see the value of the cash flows converted into their own currency decline as well.

19.3.10 Stable Yield

A consequence of the low operating costs and stable cash flows is the ability of mature infrastructure businesses to support relatively high dividend yields, typically in the mid to high single digits per annum, in conjunction with moderate capital appreciation. This contrasts with private equity or venture capital investments, which are focused on capital growth for the great majority of the investment return.

19.3.11 Low Correlation with Other Asset Classes

Infrastructure businesses typically display low correlation with traditional asset classes. One reason for this apparent low correlation is that valuation may be appraisal based, which leads to smoothed returns. When returns are smoothed, their volatilities are artificially reduced and their correlations with traditional asset classes are moved toward zero. The other reason is that infrastructure businesses are regulated and have inelastic demand, leading to stable cash flows. Therefore, while infrastructure investments are expected to provide significant diversification benefits when added to a portfolio consisting of traditional asset classes, empirical estimates of these benefits could be overstated.

19.3.12 Attractive Risk-Adjusted Returns

Total returns from mature infrastructure investments tend to range from the low to mid-teens, depending on various factors, including the sector and jurisdiction in which the business operates. For example, with exceptions, a mature regulated gas utility may generally be expected to provide a lower return to investors (due to the lower risk as a result of the regulated nature of the business and the reliability of future cash flows) compared with returns from investments in sectors such as airports and ports (in part due to the variability of revenue as a result of exogenous factors).

* * *

These fundamental asset characteristics result in a unique combination of attributes for infrastructure equity as an investment class, which makes it particularly attractive for a wide range of investors, especially those seeking to add long-dated, stable, income-generating investments to their portfolios. Infrastructure also has a number of risks associated with it, which should be mitigated where possible and reflected in return requirements when mitigation is not possible.

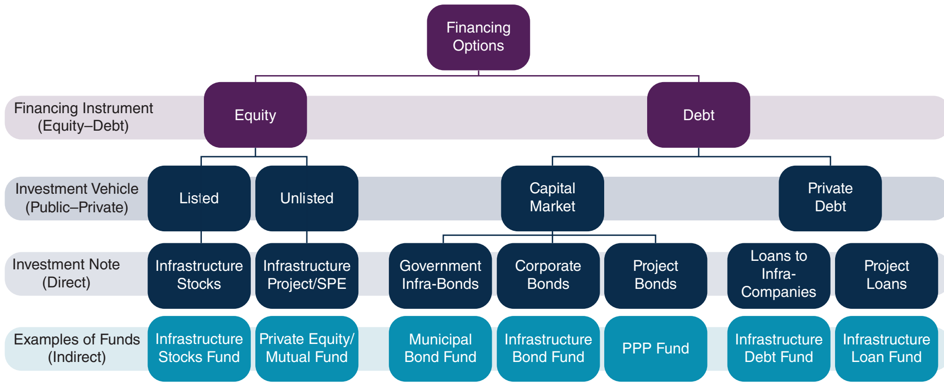
Naturally, revenue and net cash flow projections are at the heart of a commercially successful infrastructure project. These projections are a function of, among others, accurate demand forecasts, predicted future rates and tariffs, and inflation assumptions. A deep understanding of the dynamics of these variables is a prerequisite for successful infrastructure investing.

19.4 ACCESSING INFRASTRUCTURE INVESTMENT OPPORTUNITIES

Private participation in greenfield infrastructure projects is often led by construction companies and operators, which complement their own equity with debt from project finance banks and also look to dedicated infrastructure funds for additional equity. Mature infrastructure assets are often completely owned by financial investors.

There are various types of financial investors that seek to invest in infrastructure at the asset level. These include not only such institutions as insurance companies, pension funds, endowments, sovereign wealth funds, and ultra high-net-worth family offices, but also (1) infrastructure funds formed by banks and private sponsors that, in addition to their own capital, may raise third-party funds; (2) strategic buyers consisting of entities that may participate in one or more aspects of greenfield or brownfield stages, and for whom asset acquisition is a complementary strategy to spur growth; and (3) buyout-type financial sponsors with customarily higher expected internal rate of return (IRR) targets, although these investors are becoming less prevalent in the infrastructure investment space as the sector continues to become more competitive.

EXHIBIT 19.4 Infrastructure Financing and Investment Options
Source: Inderst and Stewart (2015).



There may be a degree of overlap among these investor groups. Methods for investors to gain exposure to infrastructure equity, and their relative merits, are discussed next.

Private finance for infrastructure can come in different forms and vehicles. Exhibit 19.4 displays an overview of infrastructure financing and investment options, including the main form of financing (equity and debt), investment vehicles (e.g., publicly listed and private/unlisted), and the various direct investment routes (e.g., infrastructure stocks, corporate or project bonds, and direct loans). It also gives examples of the indirect route via fund structures.

19.4.1 Equity: Private Funds

There has been increased activity in setting up dedicated infrastructure funds. Fund capital is continuously being raised to buy equity stakes in utility companies and energy, water, and transportation assets. Management teams with extensive backgrounds in infrastructure equity investment and asset management, project finance, mergers and acquisitions, investment banking, and industry help run these funds. These funds appear to be less volatile than their public market equivalents; they have private-equity-type redemption, participation, and management fee terms, and may offer access to active managers. While these private funds provide access to the potential benefits offered by infrastructure businesses, they also have some of the disadvantages of private equity investments in that they are illiquid, lack transparency, and have relatively high fee levels. Private infrastructure funds display many similarities to private equity buyout funds, but they do differ in certain important respects. The most notable difference lies in current income distribution and relatively protracted capital return schedules, given their long-dated cash flows.

Similar to private equity, private infrastructure funds are typically closed-end funds with finite lives of around 10 to 15 years. The underlying assets are typically longer-term holdings that generate cash flows and current income over more than 20 years. Capital gains typically only occur upon the partial or complete disposition of these holdings, though they may also be realized during the fund's lifetime through refinancing. The terms for privately placed infrastructure funds closely mirror those for private equity buyout funds.

19.4.2 Equity: Publicly Traded Funds

Listed funds are more liquid than private equity funds and have indefinite lives. Historically, publicly traded infrastructure funds have often specialized in specific subsectors of infrastructure. However, some recent funds have begun to invest across diversified infrastructure sectors and geographies. Although the investment, legal, and governance structures of publicly traded funds differ significantly from those of private funds, the management objectives are largely similar: to build a diversified portfolio of infrastructure equity investments that provide investors with the attractive investment attributes associated with the asset class.

However, there are certain disadvantages to publicly traded funds. As for any publicly listed stock, the price may be potentially volatile, as there are spillover effects from other sectors of the economy; therefore, returns will have a greater correlation with other asset classes, reducing the potential diversification benefits. Also, a publicly traded fund cannot easily call capital from investors when it is needed to make an investment; instead, these funds must raise excess cash (resulting in dividend yield drag) or execute follow-on issues to fund new investments (meaning that bids must be contingent on a successful equity issuance). Publicly traded funds are also subject to the same fund-specific risks as private funds.

19.4.3 Equity: Direct Deals

Direct equity investing in infrastructure assets requires an understanding of the asset, its operation, regulatory regimes, and financial structures. While dispensing with management fees, it also requires considerable experience in sourcing, screening, selecting, structuring, bidding, negotiating, and managing infrastructure assets. Many institutional investors do not possess the management skills, resources, and expertise to make direct investments in infrastructure through acquiring equity stakes or entering into joint ventures with construction companies and project operators; thus, it may not be their core competency as asset allocators. There is also significant cost to attracting and retaining a team of suitably qualified high-caliber individuals. Nevertheless, some long-standing infrastructure investors, as well as new investors, have assembled qualified and talented infrastructure teams. This skill set, coupled with resources to make significant allocations, makes them very competitive direct investors in infrastructure deals globally.

Negatives are that direct investments require very large outlays, create deal concentration risk, and reduce diversification. There is also the consideration of success rates. Unless an institution builds a large high-quality team or has exclusive relationships with leading partners, it may only be able to bid for a small number of transactions each year, which has the effect of slowing down the effective pace of investing. A number of leading infrastructure institutions complement a direct investment strategy with investment in external funds.

19.4.4 Publicly Traded Infrastructure Companies

For pure-play infrastructure investment, there is a subset of quoted stocks for investors to choose from, including certain utilities, airport companies, and toll road companies. Furthermore, the traditional preferences of the public market (e.g., low

EXHIBIT 19.5 Types of Infrastructure Investment

	Unlisted Direct Investment	Unlisted Fund	Listed Securities
Investment size	Very high	Moderate	Low
Ease of access	Difficult; takes considerable time	Moderate difficulty; takes moderate time	Easy and immediate
Length of investment	Long term	Long term	Flexible
Capacity	Limited	Moderate	High
Liquidity	Low	Low to medium	High
Leverage	Low but varies	High	Moderate
Fees/expenses	No fees; very high expenses	High fees; low expenses	Low fees and expenses
Diversification	Concentrated risk; low correlation to other assets	Medium to high; low correlation to other assets	High; high correlation to other assets
Control	Maximum control over assets	High level of control	Limited control

leverage and steady, as opposed to maximized, dividends) mean that publicly traded companies are not necessarily managed to provide the desirable return characteristics of infrastructure businesses described earlier. Further, if the illiquidity premium is one source of return to infrastructure investments, then these publicly traded investments will not match the raw return provided by private infrastructure investments.

Exhibit 19.5 provides a summary of the various options for investing in infrastructure.

19.4.5 Debt

Because the cash flows of infrastructure projects are stable and have a long duration, infrastructure projects can support a significant use of debt. The type of debt and its characteristics must reflect the underlying assets. Regulatory consideration has to be taken into account in determining the amount of leverage that can be employed. Leverage ratios for infrastructure can vary from 60% on projects exposed to high usage risk to 90% on low-risk-profile PPP projects. Investors can access infrastructure-related debt instruments in a variety of ways. Historically, institutional investors have been buying bonds issued by infrastructure companies such as utilities, pipeline operators, and transportation companies. Recently, we have seen the development of funds that specialize in investing in infrastructure loans. Loans play a more important role in financing infrastructure projects than do bonds. Fund managers are now trying to expand their offerings into this area. Most of these are closed-end, private-equity-style funds set up to provide investors access to a diversified pool of infrastructure-related loans. As described in Inderst (2012), there are a number of issues that investors need to tackle before investing in such funds or even building their own portfolios of infrastructure loans:

- Experience: The first and foremost issue is the lack of knowledge and experience. Many institutional investors, including pension funds and endowments, lack the experience to invest directly in infrastructure loans. A more common route is to invest indirectly through a fund. A problem may arise here, as many consultants are inexperienced in evaluating and recommending infrastructure debt managers.

- Fund investments: Infrastructure funds are often not pure plays, as they may invest in a variety of instruments, including mezzanine and equity instruments. Similarly, the term infrastructure related is quite common, and investors may be surprised by the types of assets they end up owning.
- Heterogeneity: Infrastructure assets are very heterogeneous, with projects ranging from high-risk renewable energy projects in emerging markets to low-risk senior loans for PPP hospitals. Cash flows of assets in different countries or sectors are likely to display vastly different properties in terms of exposure to inflation and currency risks, and to political risk.
- Debt characteristics: In that underlying assets can have different characteristics, infrastructure-related debt can also have different features. These features could differ in terms of risk (high yield versus investment grade), maturity, and seniority. Some debt instruments may have inflation protection built into them, while some may provide protection against changes in interest rates. This makes positioning of the investment in the portfolio rather complex.
- Credit risk: Infrastructure cash flows tend to be stable and insensitive to business cycles. As a result, infrastructure-related debt instruments are typically investment grade. However, a large part of infrastructure loans could be below investment grade, and covenants are not easy for investors to assess.
- Cash flow and inflation linkage: While contractually fixed cash flows create relatively safe investments, they tend to expose investors to inflation risk. Therefore, it is important to understand how the cash flows of underlying assets and instruments can be protected against inflation.
- Liquidity and concentration risk: Illiquidity is a concern for investors who find it difficult to invest when there is no secondary market. Furthermore, infrastructure funds tend to be highly concentrated in a small number of assets.
- Benchmark: Institutional investors may be reluctant to make allocations to asset classes that lack transparent and investable benchmarks. Lack of benchmarks makes it difficult to determine if the investment has been a success or a failure. Also, lack of benchmarks prevents asset allocators from using historical data to determine optimal allocations to this asset class. There are no established infrastructure bond and loan indices on the market.
- Fund structure and terms: Most of these funds are closed-end funds and structured similarly to a private equity model. Investors who are familiar with private equity investments should not assume that the terms will be identical to those of private equity funds. Fees of the fund should be closely examined, as structures are not uniform among funds.

19.5 CLASSIFYING INFRASTRUCTURE FUND STRATEGIES

Within the broad umbrella of infrastructure funds, various managers operate a range of investment strategies. These strategies may provide significant variation of investment characteristics. Therefore, investing strategy should be a key consideration for investors when assessing where an infrastructure fund belongs within an alternative investment portfolio.

19.5.1 Active Management

Some fund managers seek to make investments where their stakes afford them control or joint control. Such active managers will work closely with their portfolio companies on a range of strategic, financial, and operational issues, providing expertise from within the manager's organization and holding significant influence over day-to-day decision making. These managers often seek the ability to appoint the CEOs of the asset companies as a means of ensuring that their strategy is implemented. In order to achieve controlling stakes, active managers may have to invest significant amounts, leading to fewer but larger investments in their portfolios.

Active management in infrastructure has many similarities to the management approach of buyout funds. However, for the most part, infrastructure funds seek companies that require incremental improvement, rather than the sweeping strategic changes that typically attract buyout funds.

19.5.2 Passive Management

Other managers are content to hold minority stakes that leave another investor or investors (or government shareholders) with control. Such passive managers monitor their portfolio companies closely and offer input on certain areas (e.g., financing) that may add significant value. However, to a large extent they allow others to manage the assets, and they have less influence in decision making. Active managers typically require representation at the asset company board in order to effectively monitor their investments and to ensure that their interests are not compromised. Passive managers may be able to make smaller individual investments and thus hold a larger number of investments within their portfolios.

Passive management strategies may have more in common with those in core real estate funds. Managers of infrastructure funds will rarely describe themselves as passive, although this may be implicit from their strategy and investing history.

In practice, many managers are open to active or passive investments, although they may lean more in one direction than the other.

19.5.3 Asset Maturity

Greenfield investments in infrastructure are new developments in which no previous facilities exist. Investors should target higher returns for greenfield investments, being riskier, although they may find themselves in competition with construction contractors that may be prepared to accept lower equity returns as a means of securing lucrative construction (and sometimes operating) contracts. Brownfield investments, in contrast, are investments in facilities that already exist, although they may require additional capital for upgrading or expansion. These investments typically include a cash flow component from the currently operating facilities, which, together with some operating and usage history, makes them less risky. Depending on circumstances, both can be considered as development-stage infrastructure.

Completed assets may represent growth opportunities, either because operations have only recently begun or because they have historically been managed suboptimally. Alternatively, assets may be well established with more limited growth, meaning they are mature or in a steady state.

The risk-return profile of an infrastructure project, whether it is greenfield, brownfield, in the growth phase, or mature (with no significant capital expenditure needed), varies greatly and is manifested through its several stages. For instance, early-stage projects face construction, demand, regulatory, public acceptance, and project financing risks. Value creation for these projects with greenfield risks stems from asset value growth or from opportunities to refinance at lower interest rates.

19.5.4 Geographic Scope

Infrastructure funds may specialize in certain geographies. Different geographies may provide different returns during the economic cycle to the extent that the timing and amplitude of economic cycles vary between regions and countries. Global infrastructure portfolios, including emerging market investments, generally provide opportunities to achieve higher returns to compensate for the inherent higher risks in emerging markets. Whether these additional returns will be sufficient compensation for the added risk, particularly as emerging markets become more competitive, is debatable. There are also funds that invest only in developed countries (variously defined as OECD countries, investment-grade countries, or Europe plus North America and Australasia).

Some managers have elected to raise a number of regional funds instead of a single global fund, and there are also local fund managers offering regional or country-specific funds. These managers may be able to offer better local knowledge and government relationships. However, they have to invest in a more constrained-opportunity universe, which may lead them to be less selective if they struggle to invest their funds.

19.5.5 Sector Scope

Most infrastructure funds target a range of infrastructure sectors, largely consistent with the set described previously, potentially with a preference for economic infrastructure. However, there are also funds that specialize in utilities such as water, or transportation such as roads or airports. Specialist sector funds have advantages and disadvantages similar to those of geographic specialists: potentially more specific sector knowledge but a more limited universe of opportunity.

One area that has a large number of sector specialist funds is social infrastructure/PPPs. In particular, there are a number of well-established fund managers in the UK that specialize in private finance initiative (PFI) projects, a form of PPP used in the UK. The majority of PFI projects are for accommodation-type facilities. They do not feature demand risk or other risks that are passed through to construction and service contractors. Consequently, equity in PFI projects is particularly low risk, even for infrastructure, so returns are often at the lower end of the range (potentially single-digit IRRs).

This area is one of the most established areas of financial investor involvement in infrastructure and has become very competitive, resulting in a further reduction in returns available in recent years.

19.5.6 Core Infrastructure and Peripheral Infrastructure

A more recent development in the infrastructure fund space is the emergence of funds with relatively wide investment criteria. For example, some infrastructure funds have invested in logistics and support services companies, which may benefit from the global supply/demand imbalance for infrastructure but do not have such high barriers to entry and revenue assurance. Consequently, these assets bear more risk and should offer higher returns to compensate. This strategy may appeal to investors seeking a more buyout-style risk-return profile.

19.6 COMPARISON OF INFRASTRUCTURE WITH OTHER ASSETS

Infrastructure investment encompasses multiple sectors, geographies, stages, and investment instruments/products, each with its own risk-return attributes. This broad range makes it a versatile asset, permitting its consideration in diverse tailored portfolios. However, this heterogeneity also makes it difficult to make specific statements about the risk-return properties of this asset class. Infrastructure's capital gain potential can be viewed as an equity option attached to its bondlike, steady, and predictable return, not unlike that of a convertible bond. While infrastructure has been compared to core real estate and to private equity buyouts, it is arguably more appropriately viewed as a distinct asset class of its own.

19.6.1 Bonds versus Infrastructure

Infrastructure funds provide cash flow streams analogous to coupon payments from bonds. Since infrastructure projects are heavily debt financed, an increase in interest rates may result in higher debt financing costs that reduce earnings and distributions. In this regard, they are adversely affected, much as bonds are, by increases in interest rates (although the inflation correlation of infrastructure provides some insulation from nominal interest rate changes). However, unlike long-only bond portfolios, infrastructure project financing tends to be interest rate hedged, thereby mitigating the impact of an interest rate increase. In addition, these securities are generally not callable and hence are less negatively convex.

19.6.2 Real Estate versus Infrastructure

Infrastructure is sometimes viewed as a close proxy, though certainly not a perfect substitute, for real estate assets. The stable cash flows generated by mature infrastructure assets, such as electricity transmission cables, roads, and oil pipelines, have some similarities to the rental income streams in core real estate. However, infrastructure businesses typically feature more potential for upside and downside variation than the fixed rentals in real estate. Conversely, infrastructure greenfield projects, such as constructing new power plants, new airports, or maritime facilities, have a distinct developmental component that bears some analogy to opportunistic real estate. At the development stage, real estate can often be riskier than infrastructure, depending on asset-specific risks and the degree of mitigation. However, the differences from real estate are often missed. Not surprisingly, publicly traded, pure-play infrastructure assets have tended to be undervalued compared to real estate investment trusts (REITs) and real estate operating companies. This differential is expected to decline as infrastructure is better understood as an investment category.

19.6.3 Buyouts versus Infrastructure

Private equity buyout funds generate returns by enhancing underlying portfolio asset values via employing innovative financing terms, management changes, financial engineering, and improvements in operating efficiency. While many of these features, such as operational improvement and optimization of capital structure, are common to buyouts and infrastructure, traditional buyouts typically involve more radical changes to their target businesses in order to generate value.

Buyout funds tend to make most of their returns from capital gains generated by changes to their portfolio companies, and exit within four to six years. Infrastructure funds, in contrast, are long-term oriented, catering to investors that seek exposure to the underlying portfolio for 20 or 30 years, or even indefinitely. Value is generated from annual income and through incremental operational improvements. For funds that specialize in development-stage assets, a sale at an enhanced valuation level reflects reduced risk upon construction completion.

19.6.4 Equities versus Infrastructure

The performance of private infrastructure assets is less influenced by fluctuating demand and market cycles than by equity investments in elastic goods. Therefore, they tend to exhibit low beta to the overall equity markets, and their returns are weakly correlated with those of general equities. While some of the apparent low correlation of unlisted infrastructure investments with other asset classes might be due to price smoothing, the true correlation is likely to be relatively low due to the relative stability of their cash flows.

Exhibit 19.6 provides a summary of the comparison between infrastructure assets and a selected group of other assets.

19.7 PUBLIC-PRIVATE PARTNERSHIPS

Public-private partnerships are collaborations between public bodies, governments, and the private sector. They extend across a variety of sectors, which can include transportation, water supply, and waste management, as well as the building and managing of hospitals, schools, public housing, and prisons. PPPs can take a wide variety of forms, with varying involvement of the private sector and varying degrees of risk transfer from government to the private sector.

EXHIBIT 19.6 Infrastructure versus Other Asset Classes

Type	Financing	Operating Costs	Return	Others
Infrastructure	Long-term with leverage	Medium	Medium	Long duration and inflation protection
Real estate	Long-term with leverage	Medium	Medium	Long duration and inflation protection
Private equity	Medium-term with some leverage	Low	High	Exposure to equity risks; high leverage and interest rate risk for buyouts
Public equity	Highly liquid	Very low	High	Volatile and limited protection against inflation
Public debt	Flexible maturity and varied liquidity	Very low	Low to medium	No protection against inflation and exposure to interest rates

PPPs and regulated assets (as in economic regulation with the regulator setting prices) are normally mutually exclusive. A PPP is a finite concession with a price mechanism agreed to contractually up front. Privatized assets are owned by the private sector under a license, and prices are periodically set (or at least reviewed) by a regulator to ensure that they are appropriate and fair to the owner and to customers.

In order to enter into a PPP contract to design, build, operate, and own an infrastructure asset under a long-term government concession, private-sector investors form a special purpose vehicle (SPV) to act as concessionaire. Based on the concession contract and projected revenues from usage fees, the SPV receives equity capital from investors and enters into a contract with a design and construction company to build the asset. It may also enter into a separate contract with an operating company for the operation and maintenance of the infrastructure asset. The SPV also raises debt that can range from 60% to 90% of total project costs. Upon completion and commissioning of the infrastructure asset, the revenue generated through usage fees after payment of the concessionaire at the agreed-upon tariff rate is used to partly pay down debt, with the surplus passed on to equity contributors. PPP and PFI participation in both the development of new greenfield projects and the development of existing public-sector assets affords several benefits. In addition to providing new sources to finance capital-intensive projects, competition improves the supply and quality of infrastructure services by applying private-sector management skills, technical expertise, and efficiencies. Properly implemented, PPPs and PFIs transfer risk from the public sector to entities that may be better qualified to bear it and that are better qualified to reduce costs and create better service-delivery outcomes. It also frees government resources and public funds to address pressing social problems or to develop projects that are less attractive to private investors.

Competition for choice infrastructure assets remains considerable. Success for the qualifying bidder is conditioned, among other things, on the bidder's reputation and demonstrable ability to successfully execute the functions pertinent to the specific asset. These functions are dependent on the specifics of the stage of the asset and may include design, development, construction, financing, operating, and/or maintenance.

As with privatizations, PPPs and PFIs have faced their share of opposition from advocacy groups and, sometimes, from the public. Criticism often takes the form of allegations of reduced accountability and the profit motive of the private sector, which may be at odds with the public interest. This criticism tends to be more intense in projects in which fees charged to the public are highly visible, such as in the case of toll roads. There have also been instances (although rare) of privatization reversals in response to allegations of underinvestment and poor asset upkeep. Many issues are driven by public perception. For example, municipalities may be reticent to declare future toll increases and may have minimal incentives to publish aggressive projections, for these can be politically unpalatable. Private concessionaires, in contrast, are less reluctant to establish explicit future increases. Unlike municipal entities that borrow to meet a set capital need and thereby defer the debt problem to be addressed at a later time, private parties strive to optimize capital structures and maximize an internal rate of return. On balance, however, the vital role of PPPs and PFIs is being increasingly recognized. With government agencies strapped for financial resources, PPPs and PFIs are indeed poised to increase in importance over the coming years.

19.8 INFRASTRUCTURE REGULATION AND PUBLIC POLICY

Governments have long recognized that investments in infrastructure produce positive externalities, contribute to economic growth, expand market access, and reduce inefficiencies. From a public policy perspective, improvements in regulation, governance, transparency, ethics, property rights, and external checks are necessary and go a long way in encouraging private-sector participation in infrastructure projects. Ideally, regulatory public policy with respect to PPPs would be consistent with technological innovation, capital needs, market developments, public opinion, and changing needs to create a balance between the multiple and often opposing constituencies. Issues that inevitably feature in policy discussions relate to economic transfers and distributional effects: for example, who benefits and at what cost to others, rents, accountability, regional development, jobs, prices, and tariffs.

Given the extremely high up-front capital costs in establishing projects, infrastructure assets exhibit natural monopoly characteristics; often a single firm can produce essential social service outputs at greater efficiencies and at lower social costs than can multiple competitive firms. Recognizing this, governments restrict competition and create statutory monopolies. To preclude excess rent-seeking monopolistic behavior, infrastructure services may be procured under PPP with an agreed pricing mechanism or may be heavily regulated. In some cases, regulation has not kept pace with the swift pace of innovation in this arena.

19.9 INFRASTRUCTURE HISTORICAL PERFORMANCE

Exhibit 19.7 displays the historical performance of the Preqin Infrastructure Index along with those of some traditional asset classes and the Center for International

EXHIBIT 19.7 Performance of Infrastructure and a Select Group of Other Asset Classes

U.S. and Global Stock and Bond Indices (1Q2007–3Q2014)	Preqin Infrastructure Index	S&P 500 Index	Barclays Capital U.S. Bond Index	MSCI World Equity Index	JPM Global Bond Index	CISDM Hedge Fund Index	CISDM Managed Futures Index
Annualized total return	7.8%	8.5%	4.7%	5.4%	4.2%	4.2%	5.6%
Annualized standard deviation	9.1%	19.0%	3.3%	20.6%	6.3%	9.4%	7.6%
Sharpe ratio	0.78	0.4	1.2	0.2	0.6	0.4	0.7
Maximum drawdown	–23.7%	–44.5%	–2.8%	–48.3%	–6.6%	–19.8%	–9.2%
Correlation with Preqin	1	0.00	–0.09	–0.06	–0.04	–0.26	0.15

Source: Preqin, Bloomberg.

Securities and Derivatives Markets (CISDM) hedge fund and managed futures indices. From the first quarter of 2007 to the third quarter of 2014, infrastructure performed better than bond, hedge fund, and managed futures indices. In addition, its standard deviation was relatively low and comparable to that of the hedge fund index. While the return figures are highly time dependent, the volatility figures tend to be more stable. As has been mentioned, the relatively low volatility of return of infrastructure investments is due to the inherent stability of the cash flows of the underlying assets and some degree of data smoothing. It can be seen that infrastructure returns had virtually no correlation with other asset classes, as the reported figures are not statistically different from zero. Again, the reported correlation figures are low for the same reasons that the reported volatility figure is low (i.e., stable cash flows and data smoothing). It is reasonable to conclude that infrastructure investments have performed well in the past and have provided significant diversification benefits.

19.10 CONCLUSION

Infrastructure has traditionally been under the purview of governmental bodies. Its importance from a public interest perspective cannot be overstated. For several reasons, there has been globally widespread historical underinvestment, resulting in degradation of existing assets with a simultaneous failure to add sufficient new capacity. As this issue is unlikely to reverse course any time soon, forward-thinking countries and government entities have sought to encourage the convergence of public- and private-sector activity. Their aim is to help the creation of economically viable new infrastructure assets and upgrade, properly maintain, manage, and operate existing ones.

Formal contracting structures through active PPPs supported by recent government initiatives have made progress in encouraging private capital investment in infrastructure. However, there remains a lot to be done. Despite where they stand in the different stages of the privatization process, the collective market opportunity in developed countries remains substantial. These opportunities are in civil aviation, bridges, roads, mass transit, railways, greenfield and brownfield projects, dams, water, waste management, and energy.

The recent success of many new fund launches exclusively dedicated to infrastructure investing is suggestive of the vital role that private finance will continue to play in infrastructure development. This overall positive sentiment notwithstanding, a majority of institutional investors have yet to ramp up their investment capabilities in this area. Many of them have not made dedicated allocations to infrastructure, perhaps reflecting the relative infancy of this asset class. Consequently, possibilities for early-mover advantages remain in developed countries and are arguably even higher elsewhere. Government policy, changing legislation and regulatory norms, and private capital flows will strongly influence the evolution of infrastructure investments.

From an alternative investment perspective, infrastructure assets exhibit different risk-return profiles from other private investments; they can be a source of alpha as well as a valuable diversification tool in portfolio construction. They provide exposure through both illiquid long-term private vehicles and liquid publicly traded funds. This versatility makes them extremely attractive for various investor segments.

REFERENCES

Af2i and J.P. Morgan Asset Management. 2011. “Guide to Infrastructure Investing.” March.

Inderst, Georg. 2012. “Infrastructure Debt: A Familiar Story.” Special issue, *IP Real Estate* (Spring): 44.

Inderst, Georg, and Fiona Stewart. 2015. “Institutional Investors and Infrastructure.” March. www.georginderst.com/resources/March-2015-I3_Inderst.pdf.

OECD. 2008. “Infrastructure to 2030.” Organization for Economic Cooperation and Development, January. www.oecdbookshop.org/get-it.php?REF=5KZSL1R4XCJC&TYPE=browse.

World Economic Forum. 2014. “Infrastructure Investment Policy Blueprint.” February. www.weforum.org/reports/infrastructure-investment-policy-blueprint.

CHAPTER 20

Farmland and Timber Investments

Institutional investors have expressed an increased interest in the returns produced by the direct ownership of real assets, including farmland assets in particular.[1]

20.1 MOTIVATIONS FOR AND CHARACTERISTICS OF FARMLAND INVESTMENT

Global private investment in farmland by financial investors is estimated to be between $22 billion and $24 billion (FAO 2013).

20.1.1 Three Motivations of Farmland Investment

The rationale for such investment has typically centered around three motivations:

1. Farmland as an inflation hedge. As a real asset linked to food and energy production, farmland is expected to be a hedge against inflation. Its supply is largely inelastic (in contrast with fiat currency, securities, etc.), and increasing valuations will lead to relatively marginal increases in supply, further reinforcing its value as an inflation hedge.
2. Farmland as a diversifying source of return. Being a private market investment subject to its own physical and economic dynamics, and an asset largely held privately and often indirectly stabilized by government subsidy, farmland's returns are not, in the short run, directly linked to financial markets. Furthermore, farmland is generally a relatively unlevered asset, further disassociating its returns from financial markets.
3. Farmland as an asset positioned for scarcity themes related to food, water, and energy. Economic and demographic growth is likely to create demand for agricultural products outstripping current productive capacity and leading to the development of new farmland and price appreciation of existing farmland.

While there exists disparate research on each of these themes, empirical research is being actively conducted to address these questions from the perspective of a capital markets investor.

20.1.2 Three Characteristics of U.S. Farmland Investment

Most habitable regions of the earth have farmable land in some form; however, this chapter presents an analysis of some of the key characteristics of U.S. farmland to illustrate farmland as a global asset class. This focus is driven by three considerations:

1. U.S. farmland is a relatively stable, mature asset, and its history is free from wholesale disruptions in market structure, organizational form, and political economy.
2. The amount of available data on farmland and agriculture (both time series as well as depth of information) for the United States is greater than for any other country, facilitating the analysis of the long-run investment properties of the asset class.
3. The organizational form of farmland in the United States, which is largely privately held and market-based but subject to meaningful government regulation and activity, generally mirrors, in a mature form, the state of existing international markets for farmland (for example, Brazil, Australia, and, to a lesser extent, Eastern Europe).

20.1.3 Overview of Non-U.S. Farmland Investment

This background is essential for demonstrating the relatively stable properties of farmland, including its relationship to the macroeconomic factors of concern to institutional investors. Investment opportunities for institutional investment in farmland outside of these regions have tended to be limited. For example, Kay, Peuch, and Franco (2015) report little investment in European farmland by financial players except in Commonwealth of Independent States (CIS) countries (former Soviet republics). The Food and Agriculture Organization (FAO 2013) estimates that 10% of global farmland-related investment by financial organizations has been in European Bank for Reconstruction and Development (EBRD) countries and 6% in Africa, with the balance in the core regions of North America, Latin America, and Australia/New Zealand.

The empirical results obtained are generally supportive of recent efforts by global institutional investors to gain exposure to farmland via direct investment. The results are also relevant to other economic actors (to name a few: agribusiness, policy makers, and of course farmers themselves) seeking to understand some of the wider risk characteristics of broad-based farmland investments both in the United States and internationally.

However, investment opportunities in farmland are evolving as new risks are arising. In particular, potential risks may develop because of increasing integration (which may be expressed both in price levels and in risk correlations) between agricultural and energy markets, as recent periods of elevated fossil fuel prices have made biofuel technologies periodically cost-competitive and sources of marginal demand for agricultural commodities. A risk of great concern to cross-border investors is **expropriation** where a government takes ownership and/or control of assets belonging to foreign investors, either by direct action (nationalization or forced asset transfer) or indirect action, such as discriminatory taxation or predatory regulation.

In contrast to a period of increasing returns to farmland investment (which have recently leveled off), returns to U.S. timberland investing have undergone substantial compression in the past two decades as financial investors have acquired a large portion of timber assets previously held by integrated wood products firms. Relatedly, there has been a limited volume of transactions in core U.S. timberland assets. As a result, return opportunities for future timber investments are increasingly focused on non-U.S. assets. Return prospects and risks for such non-U.S. opportunities are discussed in a subsequent section of this chapter.

20.2 GLOBAL DEMAND FOR AGRICULTURAL PRODUCTS

The case for investing in agriculture is dependent upon both secular trends in agricultural productivity and evolving demand from end users.

20.2.1 Factors Driving Agricultural Demand and Supply

The global demand for agricultural products has experienced consistent growth, and is expected to continue to do so for the foreseeable future. Global growth rates in the demand for agricultural products have been around 2.0% since the late 1960s, and are projected by the FAO to grow at around 1.5% through 2030. According to the FAO and the Organization for Economic Cooperation and Development (OECD), three primary macro factors are driving the growth in demand for agricultural products:

1. Worldwide population growth
2. Rising incomes in emerging markets leading to changing diets, including increasing consumption of animal protein
3. The growing use agricultural products in biofuels and other non-food-based end uses

The increased production required to meet this demand will come from a combination of increased supply through:

- Improved agricultural yields
- Improved agricultural infrastructure (including irrigation, transportation networks, and processing)
- Expansion and reallocation of land under cultivation

20.2.2 Population Growth, Income Growth, and Changing Food Demands

The world's population continues to rise steadily, with almost all of the growth projected to occur outside of Europe and the Americas. Long-range population forecasts have historically been more accurate than other variables that are notoriously unreliable, such as gross domestic product (GDP). For example, world population projections for the year 2000 made in the early 1970s by the United Nations Population Division were only 2.3% higher than the actual value observed 30 years after the prediction (Sohn 2007). The UN projects global population will continue to rise by approximately 11% from 6.9 billion in 2010 to 7.7 billion by 2020, and there will be a cumulative 20% rise to 8.3 billion in 2030 (United Nations Population Division 2008). It is estimated that population expansion accounts for approximately 1% annual growth in demand for agricultural products.

Global per capita GDP has risen at an average annual rate above 4.5% since the 1980s. As the global population becomes wealthier and disposable incomes rise, dietary habits tend to shift toward agriculturally more intensive food products, such as increased consumption of meat and other animal proteins, as well as higher-value horticultural crops, such as fruit, vegetables, seeds, and nuts. This demand shift, in turn, leads toward increased demand for animal feed grains (corn, soybeans, etc.), as well as land and infrastructure necessary for the production of horticultural crops. This is important in the context of land valuation, as it is supportive to the expansion of agricultural land under cultivation—primarily, if not exclusively, outside of developed countries, where land under acre-intensive grain and oilseed crop cultivation has been static or declining—as well as the consolidation of production in areas most suitable for producing high-value horticultural crops. On a calorie basis, the production of feed grains needed for livestock production requires much more land than the production of the same calories were they consumed by humans directly in plant form. A vegetarian diet uses between one-quarter and one-half of the land corresponding to a meat-rich diet (Peters, Wilkins, and Fick 2007; Vegetarian Society 2009). However, the number of calories consumed in the form of livestock products varies greatly worldwide.

According to data available to 2011, per capita consumption of meat products in Europe (76.0 kg/person-year) and North America (115.1 kg/person-year) was around six times that in the regions with the lowest consumption, Africa (18.6 kg/person-year) and Southern Asia (7.0 kg/person-year) (FAO 2015). This indicates there is significant room for growth in meat and dairy consumption as per capita incomes rise in those regions. Further, as rapid income growth has occurred, livestock product consumption per capita has risen very rapidly in East Asia, with meat consumption growing at an annualized rate of 2.9% from 1995 to 2011 (FAO 2015). The countries of East and Southeast Asia (e.g., China, Indonesia, Japan, Philippines, and Vietnam) combined represented about 2.1 billion people in 2010, or 31% of the world's population. Despite the rapid growth that has occurred, these countries could still increase their livestock consumption by an additional 70% before they reach the average per capita consumption levels of the developed world.

South Asia, mainly comprising the countries of India, Pakistan, and Bangladesh, presents significant uncertainty, with a population of around 1.55 billion in 2010. Meat consumption in the region has remained low, barely rising over the past 45 years, which may be due to religious or cultural reasons. However, the low growth may also be attributable to the persistently low GDP per capita in these countries; according to the International Monetary Fund (IMF), nominal per capita GDP in 2014 was $1,600 for India and $1,300 for Pakistan versus a growing $7,600 for China and $54,600 for the United States. While there is significant uncertainty about the magnitude of future demands in South Asia for meat, other forms of high-value agricultural products may be subject to increasing demand. For example, India is forecast to have substantial demand for imported dairy-based animal protein (Linehan et al. 2012).

Overall, shifting diets toward higher per capita meat consumption and by-products will increase demand for feed grains. The degree to which this increased demand cannot be satisfied by improved crop yields will drive investment performance while dictating increased pressure for the expansion of farmland.

20.2.3 Biofuels and Demand for Agricultural Products

The increasing use of agricultural products in the production of biofuels has become a significant factor in agricultural demand. The combination of recent periods of rising fossil fuel prices, improvements in biofuel production and distribution, and governmental regulation and support has created a broadly positive environment for biofuels (e.g., ethanol and biodiesel). As a result, global demand for biofuels has been growing, with cumulative global biofuel production increasing at an annual rate of 16% over the period 2003–13, with 43.5% and 24.2% of global biofuel production generated in the United States and Brazil, respectively (BP 2014). While 2014/15 declines in fossil fuel prices may have further impact, prior to such declines, actual growth in biofuel production had slowed in prior years, with production growth sliding to 3.1% per annum over the period 2010–13.

Underpinned by the growing focus on energy security, a need to limit the growth of greenhouse gas emissions, and a desire to support domestic agricultural producers, U.S. government policy has been a key driver in the expansion of biofuels demand. However, changes in the assessment by policy makers of the importance of addressing issues of actual or potential energy insecurity may affect renewable energy policy, and consequently the demand and potential regulatory advantage for biofuels.

Further, biofuels growth will be supported by future growth in demand for transportation fuel, their primary use. Biofuel energy use in transportation (including both passenger and freight transportation) is projected to rise from a 2% share of total transportation fuel in 2012 to 6% in 2040 (IEA 2014).

Biofuels typically use agricultural products with food value to generate usable fuels or fuel additives. The most widely grown feedstocks for biofuels are sugarcane, used for ethanol production in Brazil; rapeseed oil, used for biodiesel in Europe; and corn for ethanol and soybeans for biodiesel in the United States (Ajanovic 2011). Given that these crops have food value, biofuel production has engendered some controversy regarding its impact on food prices, particularly during periods of high commodity prices. This controversy has abated somewhat with the recently prevailing relatively low agricultural product prices. Efforts to produce biofuels from non-food agricultural products, like corn stalks and various high-biomass grasses, have met limited success to date.

The growth in biofuels usage has created additional pressure on productivity and intensification of cropland globally, though this has moderated somewhat since 2008. In the United States, a significant portion of acreage is devoted to producing corn destined for ethanol plants. In 2014–15, an estimated 5.15 billion bushels of corn will be used for feed and residual use, which is comparable to the 5.28 billion bushels that will be used for ethanol production; 1.40 billion bushels are expected to be used for food, and 1.81 billion bushels for export. Typically, a bushel of corn (56 pounds) produces 2.7 gallons of ethanol and 17 pounds of dried distillers grains (DDGs) as a by-product that may be used as animal feed. Total *net* U.S. corn and soybean acres devoted to biofuel production—after accounting for the substitution of DDGs for corn and soymeal in animal rations—are estimated to be 28.4% of all corn acres (ethanol) and 10% of all soybean acres (biodiesel) in the United States. This represents 19% of all corn and soybean acres in total (Wisner 2015).

Increasing domestic consumption of corn has driven changes in related agricultural infrastructure requirements. For example, transportation flows from corn acres to ethanol plants reduce long-haul demand for transportation of corn itself. This also results in (1) rail shipments of ethanol to blending facilities that are located apart from the traditional export terminals of the New Orleans, Louisiana, area, and (2) the containerization of DDGs, which are typically shipped from U.S. ports located on the West Coast.

20.3 ACCESSING AGRICULTURAL RETURNS

This section discusses investor approaches to accessing the returns of agricultural assets.

20.3.1 Three Primary Approaches to Accessing Agricultural Asset Returns

Capital market investors considering nonoperating investments in agricultural assets have three primary approaches to gain access:

1. Ownership of farmland to earn lease income
2. Ownership of listed equities in agricultural firms
3. Purchase of agricultural futures or related derivative instruments

Each of these approaches to investing has distinct advantages and disadvantages, and provides access to different points in the agricultural value chain.

Capital markets investors have historically accessed futures through index-based products, such as the S&P GSCI or the commodity index now known as the Bloomberg Commodity Index (formerly DJ-UBS and DJ-AIG). The S&P GSCI is a so-called first-generation commodity index that focuses on giving investors liquid exposure to near-term price appreciation or depreciation in commodities, as well as potential benefits (or losses) associated with the roll from front to next-out futures contracts. It also offers an implicit momentum-based strategy associated with an index-weighting scheme based on accumulated value. Since its inception in 1990 until the end of 2014, the S&P GSCI Agriculture sub-index has returned an average of –1.6% per annum.[2]

Investors may also access commodity-oriented returns via investment in agricultural equities. Companies with listed equities are active at all points in the value chain. Though various providers have created indices of agriculturally related firms, these indices have been in existence for a relatively short time (typically from 2000 onward), and suffer from significant survivorship bias or insufficient attention to the changing nature of the non-agriculture-related industrial activity conducted by firms. A further consideration for investors is that a significant component of returns to agriculture-related equity investments is equity market beta, which needs to be either hedged out or accepted as a significant dilution to the desired agriculturally derived risk profile. One of the best long-term equity indices focused on agricultural equities was created by Kenneth French (KF). This index is value weighted and rebalanced annually, and requires a firm to have a contemporaneous agricultural sector classification at each rebalancing point, not just a current classification. Average annualized returns to the KF index since 1950 have been 11.6%, with 24.1% volatility.

These returns compare to the S&P 500's average annualized return of 12.8% (17.3% volatility), with a correlation of 0.55 over the same time period; holding U.S. farmland directly would have produced pure, nonrental returns of 6.1% (6.6% volatility). Rental returns to land are estimated to have had a long-run average level of approximately 6% over this time period, although they have declined recently to be around 3%. Based on correlations, there are greater portfolio diversification benefits from farmland relative to agricultural equities for an investor already holding equities. Kastens (2001) estimates the total return to owning farmland over the period 1951–99 at 11.5%, against an average return of the KF agricultural index over this time period of 10.7%. Based on a sample of 2,000 Kansas farms for the period 1973–99, Kastens also calculates returns to operating farms to be 6.8% per annum, and all-in land returns are estimated at 8.9%. This compares to an average of 13.3% per annum for agricultural equities as proxied by the KF index. According to the NCREIF Annual Cropland Index, which covers institutional investment in U.S. row cropland and is discussed later in the text, annualized total returns were 11.2% for the period 1991–2014. Row crops are crops that need to be replanted each year, such as soybeans and grains, including corn and wheat. Permanent crops are crops that do not need to be replanted annually, such as tree-based crops (e.g., apples, oranges, nuts).

20.3.2 Capturing Improvements in Agricultural Yield

While equities allow investors to access returns available in the value chain associated with agricultural production, including the sale of inputs like fertilizer, machinery, and genetically modified seeds, as well as agricultural distribution for food, fuel, and feed, direct ownership of land has a number of distinct advantages as well as risks. For example, yield-enhancing technologies (while raising the costs of specific inputs), such as the genetic modification of seeds, allow operators and owners of land to capture the bulk of the economic benefit they produce, particularly outside of Europe and North America (Carpenter 2010; Klümper and Qaim 2014). Furthermore, while investing in securities or their derivatives necessarily involves exposure to the financial markets and counterparties who may be subject to systemic risk, direct ownership of farmland offers some insulation from counterparty risks that may affect a significant portion of an institutional investor's portfolio.

More generally, the long-run returns to owning farmland, and the short-run returns to owning particular types of farmland, require an understanding of crop yield. Crop production is largely a process of transforming solar energy into chemical potential energy; land, with water, is the platform upon and through which that process occurs. **Crop yield** is a measure of agricultural productivity expressed in units produced per unit of land used for a specified unit of time. Assuming (1) no scarcity of water, (2) sufficient nutrients, and (3) no temperature extremes, crop yield can be decomposed as (Hay and Porter 2006):

$$Y = S \times I \times E \times H \tag{20.1}$$

where:

Y = yield
S = total solar radiation over the area per period
I = fraction of solar radiation captured by the crop canopy
E = photosynthetic efficiency of the crop (total plant dry matter per unit of solar radiation)
H = harvest index (fraction of total dry matter that is harvestable)

S is largely a function of geography and weather. Over the short run, I is the most variable, as it depends on the extent to which crops have been able to deploy canopy; total dry matter production (absent severe stresses of drought, etc.) is largely a linear function of captured solar radiation. Increases in H have accounted for a significant increase in yields to key grains over the twentieth century, as shown in Exhibit 20.1. E varies little, but is the subject of much research. As for a given species, I and H have been changed for many key crops and are difficult to change further.

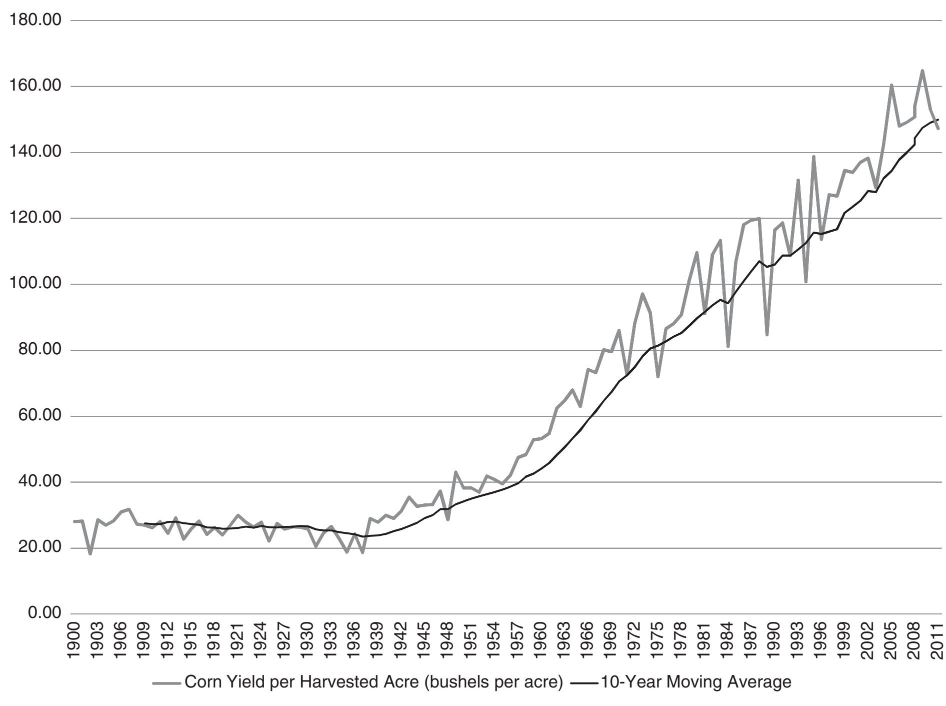


EXHIBIT 20.1 U.S. Corn Yields Are Rising over Time
Source: Author's calculations from U.S. Department of Agriculture (USDA) data.

APPLICATION 20.3.2

A parcel of farmland experiences no change in its units of total solar radiation or the percentage of solar radiation captured by the crop canopy. A new seed allows the crop to experience a 25% increase in photosynthetic efficiency but causes a 10% decline in the percentage of the total dry matter of the crop that is harvestable. What would be the change in the crop yield from using the new seed?

Using Equation 20.1, the change in the crop yield can be expressed as the product of the increase (>1) or decrease (<1) in each of the four components:

$$(Y + \Delta Y)/Y = (S + \Delta S)/S \times (I + \Delta I)/I \times (E + \Delta E)/E \times (H + \Delta H)/H$$

With $(E + \Delta E)/E = 1.25$ and $(H + \Delta H)/H = 0.90$, $(Y + \Delta Y)/Y$ must equal 1.25×0.90 or 1.125, indicating a 12.5% increase in crop yield.

Note that changes in any of the four variables on the right side of Equation 20.1 can be expressed using the change in the yield (Y) and any changes in the other three variables.

20.3.3 Factors Driving Growth in Agricultural Yields

Growth in yields, particularly in the developed world, has occurred largely as a result of four advances:

1. Improved technology (including advancement of seed stock through plant breeding and, in certain cases, transgenic modification)
2. Improved agronomy
3. Increasing use of inputs, such as fertilizer
4. Increasing use of capital assets, like machinery, and agricultural infrastructure

The role of genetic advancement in yield gains is often mentioned, but is not well understood in the practitioner literature. There are a number of important points to be made here, including:

- Duvick (2005) surveys the contribution of plant breeding to improvements in corn. He estimates 50% to 60% of yield gains from 1930 to 2001 are the result of genetic enhancement.
- Genetic enhancement has enabled yield gains associated with optimized agronomic practices. One the most important for corn is increased planting density, or the closeness of planted seeds, for which Duvick cites evidence that more than 20% of yield gains have been attributable to increasing plant density. In order for increased planting density to improve yields, corn breeders have focused on the development of corn lines able to tolerate the environmental stress resulting from within-crop competition for nutrients. However, he also notes that opportunities for yield advancement via increased density is now limited.

- While genetic advancement has been able to raise overall yield levels in corn, it has been less successful in doing so for other crops, like wheat and soybeans. Specht et al. (2014) find representative U.S. on-farm yield growth for soybeans has been 0.43 bu/acre per year since 1983 and genetic yield gains have been of a similar order of magnitude. However, this obscures ongoing changes in how plants are bred and adapt to climates (cultivar and agroclimactic production zone), and more careful analysis indicates a better estimate of the contribution of genetic gain is roughly two-thirds of total yield gains. For example, gains have been 0.62 bu/acre per year for irrigated (Nebraska) soybeans, of which 0.40 bu/acre per year is due to genetic improvement. Suhre et al. (2014) demonstrate increased plant density has also played an instrumental role in increasing soybean yields. (See Exhibit 20.1.)
- Smith et al. (2014) offer a comprehensive review of yield gains in major U.S. field crops. Outside of corn and soybeans (in such crops as alfalfa, barley, and potatoes), genetic advancement has played a limited role in increasing crop yields. Changes in yields for these crops have largely been achieved through improved agronomics.

20.3.4 Determinants of Agricultural Profitability

While crop yield is an important factor in the economics of crop production, the *price of harvested crops* is the most important and volatile short-run determinant of farming profitability.

A further consideration is the operating overhead of owning and operating a farm. The people, expertise, chemicals, machinery, seeds, livestock, and so on needed to make a farm productive create a substantial aggregate (and ongoing) cost. As an idiosyncratic physical asset, professional fees such as surveying, legal, and other administrative services may add up. Whether these costs are wrapped into a single operator's fees or are paid piecemeal by the investor, the cost of running a farm portfolio is far higher than the custody costs for a portfolio of financial assets, and must be budgeted carefully to arrive at a robust business case to validate an investment.

Moreover, it is important to recognize that merchandisers, processors, and transport and logistics providers will earn part of the value created between the farm gate and the end user. For example, the farm share of retail wheat flour is approximately 26%. Less extreme is the case for bulk commodities, like soybeans and corn. Margins between different geographic locations along the value chain are tighter, but significant. For example, the difference between the point of sale at a country elevator and a point of export, such as the Gulf of Mexico, is routinely in excess of 10% of the value of the crop. The transportation cost portion of total landed price for crops shipped from the U.S. Midwest to Asian port destinations can often be up to 25%, even during periods of relatively low ocean freight rates.

An examination of Exhibit 20.2 shows that the profits to farming row crops can be quite volatile, as the costs of both inputs and outputs are variable. If the price of corn declines at a time when the operating costs (e.g., fuel, chemicals, and fertilizer) rise, farmers will face rapidly declining margins, or even operating losses.

EXHIBIT 20.2 U.S. and Selected Regional Corn Cost-of-Production Estimates for 2001, 2005, and 2010

	Dollars per Planted Acre			
Item	Heartland	Northern Crescent	Prairie Gateway	United States
2001				
Operating Costs	154	162	198	161
Overhead and Ownership Costs	211	225	274	221
Total Production Costs	340	343	389	344
2005				
Operating Costs	185	198	219	188
Overhead and Ownership Costs	251	267	309	259
Total Production Costs	388	394	417	387
2010				
Operating Costs	295	295	273	289
Overhead and Ownership Costs	384	378	384	382
Total Production Costs	572	514	512	550

Source: U.S. Department of Agriculture.

20.3.5 Government Subsidies and Agricultural Returns

U.S. farmland has historically been the key point of focus for government-sponsored agricultural support, insurance, and stabilization payments. These payments are capitalized into the value of land, tending to raise values and providing a floor to possible price depreciation. Estimates have varied widely, and depend on the prevailing policy regime regarding payments and insurance policies. For example:

- Goodwin et al. (2003) quote studies estimating that between 7% and 69% of farmland value can be attributed to capitalized government payments, with land in the Northern Great Plains most dependent on government payments.
- Based on data from 2000 to 2006, Ifft, Kuethe, and Morehart (2015) find that certain government payments are capitalized in land prices, with a dollar of incremental payments valued at $18, or a cap rate of 5.6%. They find that estimates of the capitalized value of such payments are noisy, but Plains cropland exhibits the most statistically significant result.

However, studies such as Kirwan and Roberts (2010) and Gardner (2003), which examine the cross-sectional variation of land prices within particular regions, find land prices are not sensitive to government payments. Gardner hypothesizes this may result because subsidies do not have a commodity-specific impact on land values, since land use is highly flexible, particularly over the long time horizons associated with capitalization. Studies also suggest that land rental rates are largely set so landowners may capture such subsidies. For the analyst considering farmland valuation, best practice would be to consider these four factors:

1. The potential income or risk mitigation attributable to government subsidies
2. The parties who might be eligible to receive such payments if operators are distinct from owners
3. The impact on valuation and rental rates
4. Conducting sensitivity analysis to determine the impact of potential changes in applicable government payments

20.4 UNDERSTANDING THE RETURNS TO FARMLAND

Understanding the returns to farmland is enhanced by analysis of historical returns, macroeconomic factors, heterogeneity of the returns, and the impact of changing prices for land.

20.4.1 Historical Returns to U.S. Farmland

There are several considerations when evaluating the historical returns to U.S. farmland:

- Pre-Depression (pre-1930s) farmland returns were highly volatile during a period when there were few government programs aimed at stabilization of agricultural production and incomes.
- Exhibit 20.3 shows a strong trend emerging around 1940, starting a period of consistent annual growth averaging about 6%. Note: this growth is not a total return index, as it does not include ancillary cash flows, such as lease payments.

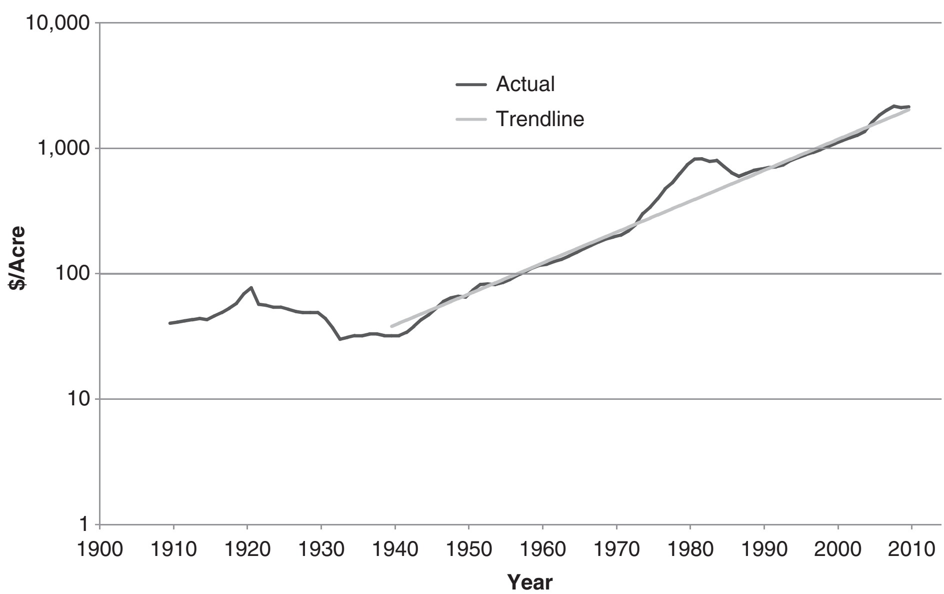


EXHIBIT 20.3 Farmland Prices Are Rising over Time
Source: Author's calculations using USDA data.

- There have been only two deviations from this trend: a strong rising trend starting in 1973 (reverting to a lower value in 1986), and a weaker higher trend beginning with the recent commodities boom that began in the early 2000s and continued until the crash in commodity prices in the early and mid-2010s. Farmland price growth in core heartland areas has recently begun to level off, and even decline, in some areas.

20.4.2 Macroeconomic Factors Explaining U.S. Farmland Returns

The returns to a specific investment in farmland should be driven by the cash flows generated on an investor's land. This cash flow is a function of microeconomic factors that are asset and sector specific (soil quality, potential yield, input costs, nearby agricultural infrastructure, etc.). Generally these microeconomic factors are a substantial determinant of cash flow; in particular, the quality and experience of management will differentiate one farm from another. Good farm management includes consideration of the long-term productivity of land with its near-term yield; for example, farmers routinely rotate crops (e.g., soybeans against corn) in order to maintain soil quality, along with limiting the cost of purchased inputs and the development of crop-specific pests and diseases. Investors acquiring and operating farmland should pay particular attention to structuring economic incentives for the operator that reflect the multiple time horizons of farmland husbandry. In addition to selecting a specific subset of farmland assets expected to generate positive returns, the investor will likely want to consider the ramifications of macroeconomic factors on the farmland investment as well as the effects on the existing portfolio. That is, a capital markets investor is keenly concerned with the accumulation of macroeconomic and asset-based factor exposures in the investment portfolio.

An analysis was conducted of the factors that play a part in driving U.S. farmland returns, as proxied by the USDA $/Acre series for the period 1973–2009. Factors considered included U.S. GDP, U.S. industrial production, U.S. inflation (as proxied by the GDP deflator or CPI-U), interest rates (as proxied by the yield to worst on the Barclays Aggregate Index and the returns to the Barclays Government Bond Index), strength of the U.S. dollar, European inflation, and the spot prices of commodities like oil, corn, and wheat.

A standard regression analysis was conducted to identify the most important factors, after which variables were selected based on significance, and the model was reestimated on the subset of relevant variables (see Exhibit 20.4). The model is statistically significant, with an adjusted R-squared of 0.73. Interestingly, the most significant variables are:

- U.S. Consumer Price Index (CPI), which shows that the returns to U.S. farmland have been a significant hedge against inflation risk.
- Yield to worst, which is an indicator of the level of interest rates with a negative coefficient. This suggests that higher interest rates are associated with lower farmland returns. This association is likely linked to both the business cycle (as higher interest rates are associated with a contraction in monetary policy) and the fact that higher interest rates are likely to put downward pressure on land

EXHIBIT 20.4 Regression Explaining the Prices of U.S. Farmland, 1973–2009

Variable	Coefficient	Standard Error	*t*-Statistic	Probability
Constant	0.076	0.019	4.033	0.000
Wheat	–0.044	0.047	–0.921	0.364
Corn	0.046	0.048	0.947	0.351
Yield to Worst	–2.549	0.387	–6.584	0.000
DXY	0.218	0.074	2.952	0.006
U.S. CPI	0.204	0.382	8.397	0.000
Industrial Production	0.695	0.199	3.494	0.002
Oil Price	0.032	0.019	1.673	0.105
R-Squared	79%			

Source: Authors' calculations based on USDA data.

prices (as increased discount rates bite into the present value of future proceeds from agricultural land).

- Industrial production, which is positive, suggesting land prices are procyclical and part of a growth story, rather than explained as part of a "store of value" dynamic.
- The U.S. Dollar Index (USDX or DXY), which suggests that a stronger dollar is associated with increases in land prices. This may be a proxy for monetary policy, or may actually reflect the impact of increased external demand for U.S. farm products on both land values and the price of the dollar.

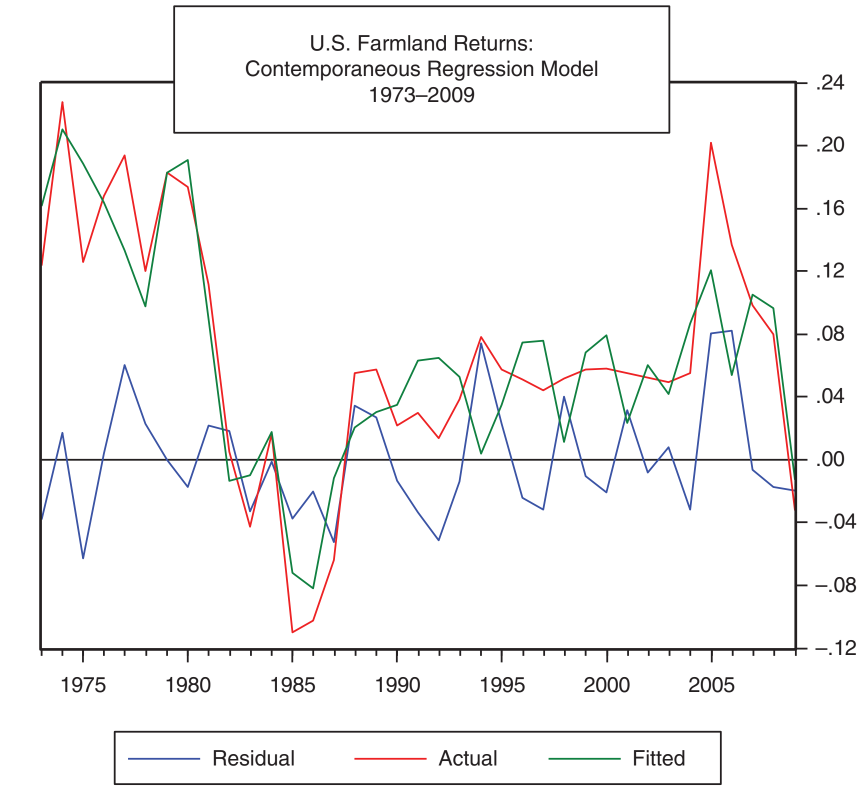
In particular, note that changes in the spot prices of commodities—corn, wheat, and oil—are not statistically significant determinants of contemporaneous land values. It is, however, expected that in the medium run elevated crop prices may have positive impacts on farmland prices, all other things being equal.

The robustness of this model can be evaluated in two ways:

1. The results predicted by the model (fitted) are compared to actual farmland returns. From an inspection of Exhibit 20.5, the model's fit has been relatively strong over the entire period, while not being driven by outliers or other factors.
2. A more formal analysis shows that the regression parameters are statistically stable, based on the cumulative sum of squared deviations from a set of recursive (Durbin) regressions.

20.4.3 Heterogeneity of U.S. Farmland Returns

The analysis presented in Exhibits 20.4 and 20.5 was conducted on an overall average of U.S. farmland returns. However, investors seeking to make investments in farmland cannot invest in such an average because, given available investment vehicles, farmland investments are typically focused in a small number of states. The question is: How heterogeneous are farmland returns across U.S. states? The simplest way to evaluate this is to conduct a principal components analysis for state-level returns for the period 1973–2009. The results suggest, first and foremost, there is a common factor (first principal component) across U.S. farmland, which explains 56% of the



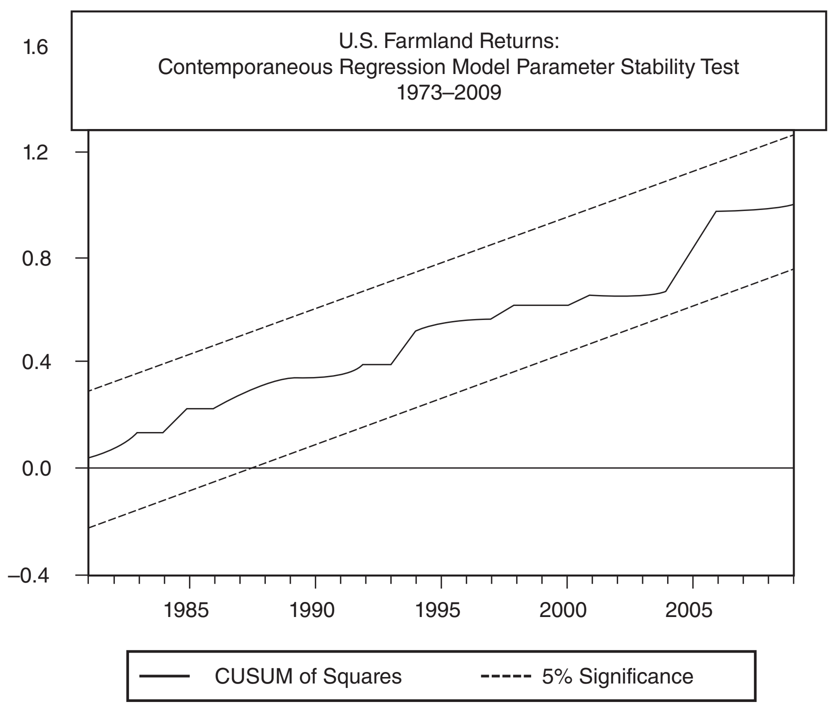


EXHIBIT 20.5 Regression Fit for U.S. Farmland Returns, 1973–2009
Source: Author's calculations from USDA data.

cross-sectional variation in farmland returns over this period. Examining the correlations between the first principal component and state-level returns, Kansas and Missouri returns are observed to be highly correlated (0.90 and 0.91, respectively) with this common factor. Based on this analysis, the risks and returns associated with Kansas and Missouri are representative of U.S. farmland returns, and vice versa. Geman and Martin (2011) provide further analysis, which groups state-level farmland returns according to their macroeconomic sensitivities. They find that coastal farmland returns have significantly different macroeconomic sensitivities than core farmland in primarily agricultural states. One major factor that tends to be present in coastal farmland, and that accounts for a portion of this difference, is urbanization and commercial or residential real estate use value. That is, the land in the U.S. Midwest used for farmland is farther away from residential development than farmland closer to the U.S. East Coast and West Coast regions. Coastal regions also tend to have higher land and housing prices in relatively close proximity to farmland. This introduces a speculative element of potential rezoning into the appraisal of the investment characteristics of agricultural land.

The preceding analysis, while helpful, is based on historical USDA data, which is survey-based and does not actually represent returns to farmland investors. The National Council of Real Estate Investment Fiduciaries (NCREIF) publishes data on the actual estimated performance (via transaction and appraisal values) of farmland investments among its members. These indices have the following characteristics:

- The data are based on reported returns (appraisal and transaction based, gross of portfolio level fees but net of individual property level fees) to institutions investing in U.S. farmland, for agricultural properties only, including permanent, row, and vegetable cropland.
- Investment returns are reported on a nonleveraged basis; while there may be properties in the index that have leverage, returns are reported as if there were no leverage.
- The properties must be owned or controlled by a qualified tax-exempt institutional investor or its designated agent.
- A specific property is included in the index beginning the first full quarter it qualifies (properties are generally excluded in the acquisition quarter).
- Only income-producing properties are included.

The Farmland Index constructed by NCREIF represented about $5.5 billion in assets as of 2014, which is about 2% of U.S. farmland value, though this is growing significantly as institutions are deploying capital to farmland. The index itself is currently split between annual row crop assets and permanent crop assets.

Row cropland is annual cropland that produces crops such as corn, cotton, carrots, or potatoes from annual seeds. Row cropland comprises approximately 55% of the NCREIF Farmland Index. **Permanent cropland** refers to land with long-term vines or trees that produce crops such as grapes, cocoa, nuts, or fruit. Permanent cropland comprises approximately 45% of the NCREIF Farmland Index. As shown in Exhibits 20.6 and 20.7, over the past 20 years, total net returns to row crop land have averaged around 11.2%, split roughly evenly between income and capital appreciation. While returns to permanent crops have been of a similar order of magnitude (12.6%), they have been comprised largely from income.

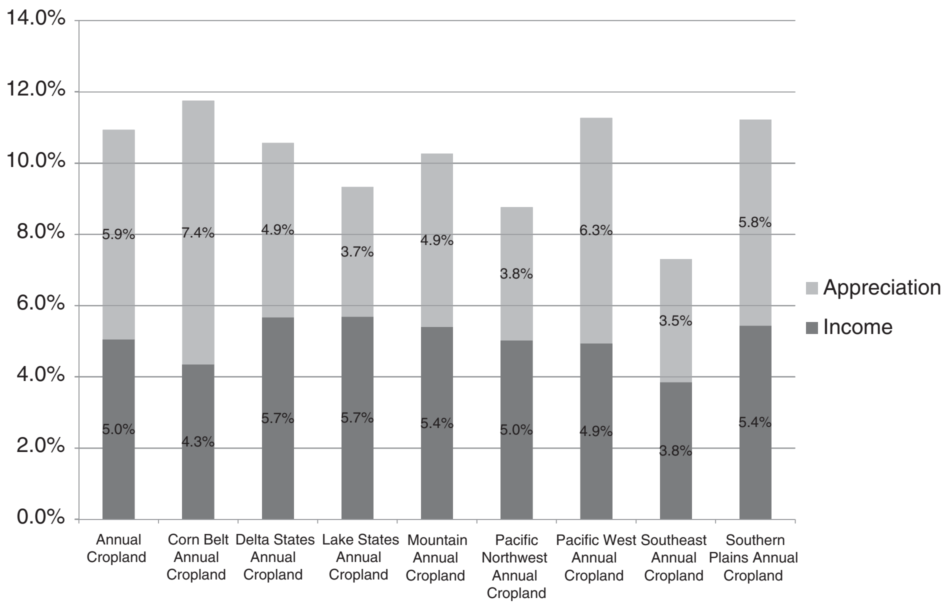


EXHIBIT 20.6 Annualized Returns to Annual Cropland, 1991–2014
Source: NCREIF.

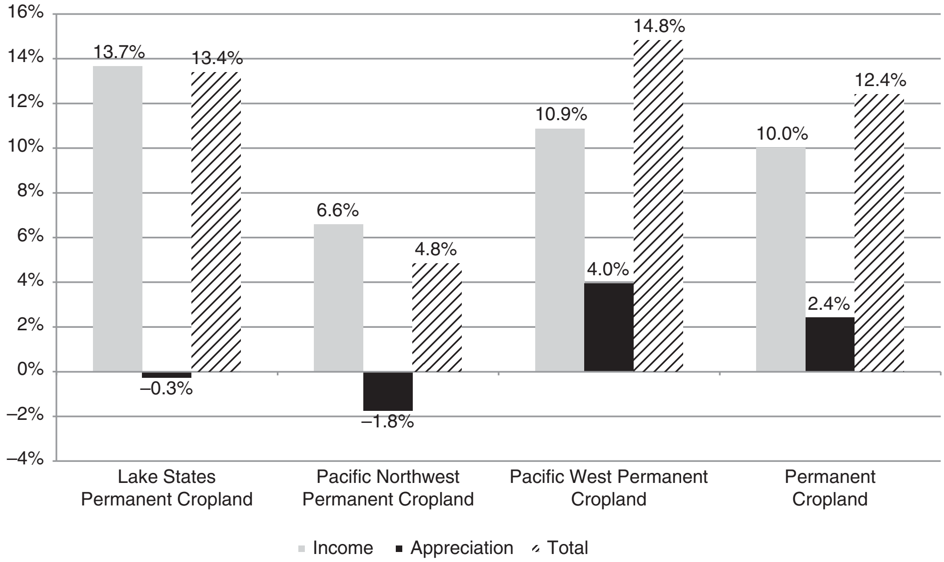


EXHIBIT 20.7 Annualized Returns to Permanent Cropland, 1991–2014
Source: NCREIF.

20.4.4 Farmland Price Appreciation and Farmland Returns

Investors contemplating commitments to farmland should consider the causes and consequences of significant price appreciation in U.S. farmland over the period 2009–14. An important public policy concern, as well as an investment management issue, is the extent to which institutional flows cause/lead price appreciation in U.S. farmland markets. Alternatively, one can also ask if institutional commitments lag changes in valuation. Martin (2014) provides evidence of forecasting potential using a Granger-causality analysis. A **Granger-causality analysis** is a hypothesis test to determine if one time series is useful in forecasting another. Results indicate that institutional flows do not lead farmland capital appreciation or income. Farmland capital appreciation does, however, lead income. Evidence of a more anecdotal nature suggests a significant driver of farmland price appreciation, particularly in heartland markets, is driven by the willingness of existing farmers to pay elevated prices for land at auction or otherwise at sale. For example, only an estimated 18% of Iowa land purchases in 2013 were made by investors; 77% were made by current farmers (Martin 2014).

20.5 INVESTING IN AGRICULTURAL INFRASTRUCTURE

Reflecting the increasing appreciation of the investment benefits of including farmland assets (Geman and Martin 2011) and private infrastructure (NBIM 2013) in institutional portfolios, institutional investors recognize the potential portfolio benefits of including direct investment exposure, whether structured as private equity (PE), publicly traded equity, or agricultural infrastructure. This is similar to the evolution in thinking that institutional investors went through with regard to energy-related investments: Investors first focused on acquiring direct exposure to the commodity through the ownership of resources and production capacity, and then expanded their exposure to include energy-related infrastructure, like pipelines. Morgan Stanley (2013) estimated the total market cap of master limited partnerships (MLPs) that focus on midstream infrastructure assets at $480 billion. Recent allocations from private infrastructure funds have focused on traditional energy and renewable energy (40%+), with U.S.-centered deals more heavily focused on traditional energy.

20.5.1 Four Economic Functions of Agricultural Infrastructure

As discussed in Wood Creek Capital Management (WCCM 2015), **agricultural infrastructure** is composed of substantial capital assets designed to enable or improve the efficiency of the production, handling, and distribution of agricultural commodities. This typically involves a capital asset that need not be specific to any particular type of agricultural commodity.

We can be more precise about the specific function of agricultural infrastructure, and in so doing classify agricultural assets according to their role in four economic functions:

1. Location. Transportation and logistics improve spatial efficiencies, allowing for improved optimization of inputs as well as demand satisfaction.
2. Time. Storage facilities allow harvested product to be delivered when most needed, as well as promote collection and buffering to improve scale efficiency in transportation and processing.
3. Quality. Blending, grading, and basic processing facilities allow for efficient matching of products with end users, as well as transformation of agricultural commodities into forms more readily transported, traded, or consumed.
4. Input allocation and delivery. Irrigation and other fixed input delivery structures improve yields, reduce production volatility, and minimize inputs per unit of output through controlled delivery (e.g., scheduled irrigation).

20.5.2 Three Drivers of Productivity of Agricultural Infrastructure

Agricultural infrastructure, like other forms of infrastructure, derives economic return largely from the value of efficiency gains. Supranormal returns may be earned through the exercise of market power.

The key economic function of agricultural infrastructure is to increase productivity of the agricultural value chain. Therefore, any broad analysis of the economic and investment benefits of agricultural infrastructure must consider the secular and changing dynamics in agricultural productivity, including these three drivers:

1. Intensification: Additional inputs leading to greater outputs
2. Extensification: Increased or decreased land use
3. Specialization: Monocropping or prescribed intertemporal crop rotation

An example of changing agricultural infrastructure is found in the changing amount of private investment devoted to grain storage facilities. Such facilities may be broadly classified into:

- On-farm storage: Smaller in size and broadly distributed geographically
- Off-farm storage: Larger in scale and integrated into the grain transportation network

USDA data show that storage capacity is growing both on-farm and off-farm. Part of the reason is attributable to increasing crop sizes, as well as efforts by farmers to improve efficiency and avoid congestion costs around harvest times through increasing on-farm storage capacity. Providers of storage to producers, as well as other players in the agricultural infrastructure value chain, often generate revenues more tightly linked to the volume of agricultural goods handled (as opposed to the price level of those goods). They are able to earn returns that are diversifying relative to direct land-based investments, which may themselves be sensitive to overall commodity price levels. Econometric analysis does not indicate whether on-farm storage changes lead or lag off-farm storage changes, or vice versa (WCCM 2015).

20.6 GLOBAL INVESTING IN TIMBERLAND[3]

While some institutional investors are recently discovering farmland investing, most have long recognized the benefits of investing in timberland assets. The motivations for timberland investing have been particularly focused on inflation hedging and diversification properties of the asset class. Institutional investment in timber assets, as proxied by the NCREIF Timberland Index, was equal to $22 billion in 2013 (Davis et al. 2014). This is well documented in a range of articles (Meketa Investment Group 2010; Healy et al. 2005; Waggle and Johnson 2009). R&A Investment Forestry (2014) reports that timber investment management organizations (TIMOs) have an estimated $45 billion of assets under management (AUM) as of 2014. Fu (2014) provides a sustained overview of the timber investment industry and considerations for investors contemplating investments in this area.

20.6.1 Four Key Attractions to Timber Investment

Timber-related assets have long held attraction for long-horizon investors seeking assets that should hold value against changes in inflation. Four of the key reasons timber has been so attractive are:

1. Inputs have relatively low cost and, therefore, limited sensitivity to inflation.
2. Products are homogeneous and used throughout the economy, and are therefore unlikely to be exposed to substantial idiosyncratic price volatility.
3. The harvesting option is a real option allowing the owner to generate cash flows at the time of elevated demand.
4. It is a long-duration asset, suitable for matching long-duration liabilities.

While there are many reasons why timber is thought to be an inflation-robust asset, there is surprisingly little published empirical evidence investigating this reasoning. A few exceptions:

- Washburn and Binkley (1993) offer an in-depth analysis of the relationship between forest assets and inflation. They analyze data based on historical “stumpage” prices rather than the entire package of timber and land revenues. Specifically, they focus on unexpected inflation in the United States for the period 1953–87. Their results indicate that Western and Southern timber assets are positively correlated with *unexpected* inflation. Maine forestry asset values are positively related to unexpected inflation, but not in a statistically significant manner. Washburn and Binkley also find that the beta of forestry assets relative to *expected* inflation is generally positive but rarely statistically significant.[4]
- Washburn, D’Anieri, and Aronow (2004) conduct a similar analysis as Washburn and Binkley, using NCREIF data that includes land and timber returns, and find similar results.
- Martin (2010) provides further analysis of the inflation-hedging properties of timber and related wood product investments. NCREIF Timberland Index investment returns are compared to an investment proxy associated with timber-only assets. Evidence is presented showing that the inflation-hedging properties of timber-related assets are more strongly linked to the timber and less to the underlying land.

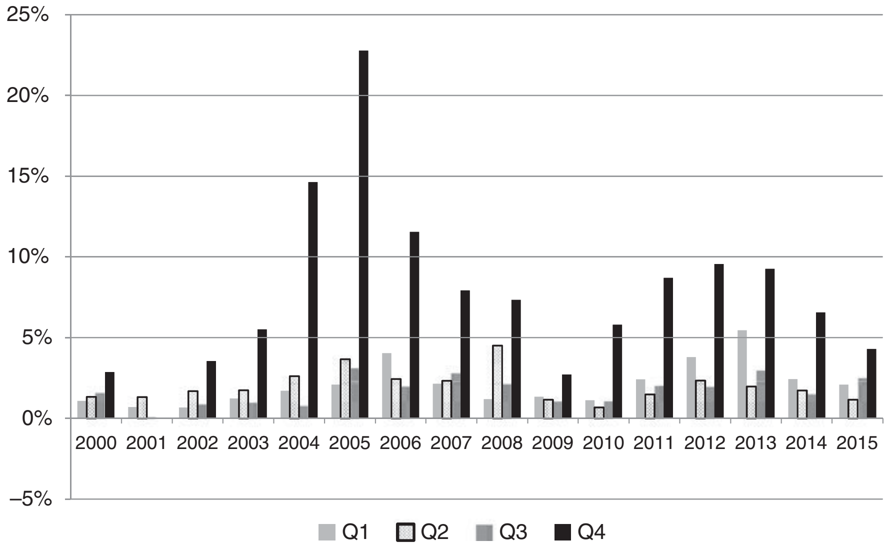


EXHIBIT 20.8 Annual Returns to Timberland Investments
Source: NCREIF.

20.6.2 Returns to Timberland Investing

While there is significant interest in the return properties of timberland, as shown in Exhibit 20.8, the 2000s witnessed substantial compression in returns to investing in U.S. timberland relative to the 1990s. Institutional investors, along with other sources of private capital, have acquired an increasing proportion of overall timber assets. These acquisitions have largely come from the prior holdings of integrated forest product companies. These companies correctly recognized that returns available from retaining forest land assets in vertically integrated corporate form would be less than those available from the divestiture of such assets to investment capital (Healy et al. 2005; Lonnstedt and Sedjo 2012). Over the period 2000–9, an estimated 40 million acres changed hands. This divestiture process is not globally uniform, and reflects both cost of capital as well as addressing market structure and regulatory issues. Canada, Finland, and Sweden have also seen some divestitures.

More recently, timberland returns have rebounded, with NCREIF Timberland Index returns averaging 13.75% per annum from 2010 to 2015. However, the recent transaction environment has been difficult. Only 1.4 million U.S. forestry acres, worth $1.4 billion, changed hands between 2010 and 2013, and an estimated $4 billion in undeployed capital remained on the books of TIMOs (R&A Investment Forestry 2014).

This compression in returns and reduction in transactions in North America has resulted in substantial attention by investors to non-U.S. timber and forest product opportunities. However, little research has appeared on the investment characteristics of these opportunities. This lack of robust, publicly available research may indicate an underexploited, and therefore potentially attractive, investment opportunity set. One notable exception: Akers and Staub (2003) present a factor-based approach to international timber returns and portfolio construction.

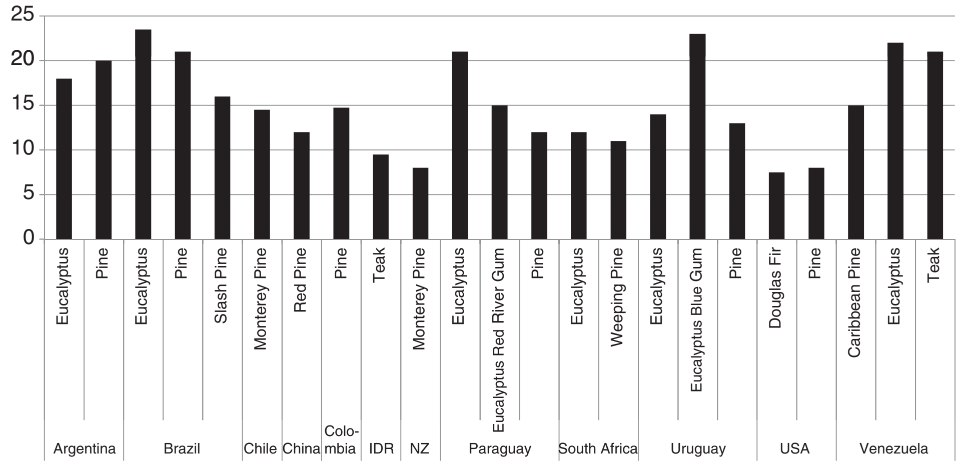


EXHIBIT 20.9 IRRs of Timber by Country and Species, 2008
Source: Cubbage et al. (2010).

Of some remark in helping investors evaluate potential returns to timber investments outside of the United States, Cubbage et al. (2010) published a comprehensive set of estimated internal rates of return (IRRs) for representative species and locations globally (Exhibit 20.9). The reported IRRs are between 5.2% to 14.6%. The reported IRRs vary greatly among countries and within each country depending on the species. These are reinvestment IRRs, which assume land costs have already been paid.[5] Of particular note, of course, are the relatively low IRRs for U.S.-based species when compared to non-U.S. species. Eucalyptus species tend to have the highest IRRs, in part because of the shorter periods to rotation (15 to 20 years).

20.6.3 Risks to International Timber Investing

While the IRRs for species grown outside of the United States are much higher than for U.S. species, the risks to an unhedged, nonlocal currency investor are substantially greater. In particular, incremental risks that must be considered include:

- Currency risk. Typically, currency volatility between developed and emerging market currencies is on the same order as asset price risk in local currency. Given the long-term nature of these investments, currency hedging can be costly, as well as difficult. It remains an open question as to whether currency risk between developed and emerging markets pairs involves compensation for risk bearing (Kim 2012).
- Legal risk. Land-based assets are typically subject to a unique, and political, set of regulations governing foreign ownership, taxation, environmental obligations, and indigenous rights. Agricultural legislation often includes some of the oldest statutes in any given jurisdiction, and can therefore be among the most idiosyncratic. In particular, plantation-based forestry may be subject to different regulations than would economic use of existing forests. Also, environmental activists have targeted several timber-related projects.

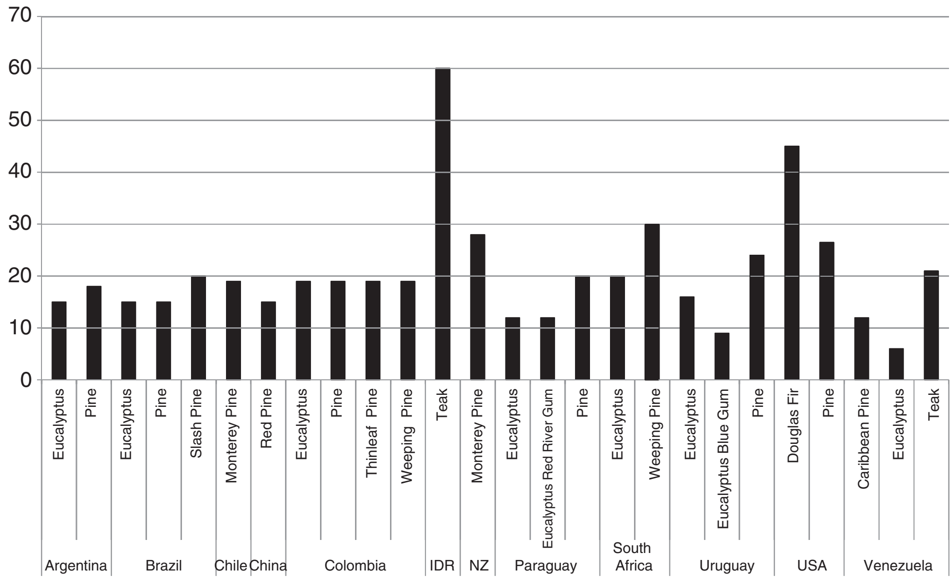


EXHIBIT 20.10 Optimal Rotation Age of Timber Varies by Species and Location
Source: Cubbage et al. (2010).

20.6.4 Rotation

Cubbage et al. (2010) show the rotation age by species. **Rotation age** is the amount of time, usually expressed in years, for timber to reach the size for economically optimal harvesting. As shown in Exhibit 20.10, most species of trees take between 10 and 20 years to reach a size that allows for economically optimal harvesting. Beyond a certain age, the growth rate of trees begins to slow. At that point, the forestland may be better used for newly planted trees, as the existing trees are harvested after the peak growth rate has passed.

20.7 FARMLAND AND TIMBER INVESTMENTS COMPARED TO OTHER REAL ASSETS

Comparing farmland and timber investments to investment in other real assets begins with a discussion of the characteristics of real assets.

20.7.1 Characteristics of Real Assets

While there is some consensus on which asset classes may be considered within the real asset universe, there is little consensus on exactly what properties of an asset class entitle it to be considered a real asset. Plan sponsors and other institutional investors have a wide range of reasons for considering a particular asset class a real asset, which typically include one or more of the following characteristics (Martin 2010):

- Inflation hedge: Positively correlated with U.S. or European price inflation
- Intrinsic value: Expected to preserve value during periods of financial market contagion or substantial changes in the economic environment due to changes in the business cycle
- Scarce input: Often benefit directly from the increasing scarcity of production inputs, particularly in core economic sectors such as energy, manufacturing, and agriculture
- Economic infrastructure: Are essential components to economic infrastructure, including the built environment (commercial and residential real estate); transportation (including roads, rail, shipping, and air); and major projects (such as telecommunications and pipelines)
- Long term: Offer long-term risk and return properties suitable for investors seeking to fund long-term liabilities.

Institutions considering investments in real assets tend to focus on a number of key investment characteristics that vary across specific real asset classes, in addition to the economic rationale(s) for the asset class being considered a real asset. See Chapter 21 on intellectual property for an explanation as to why some forms of intellectual property may be considered to be a real asset. Also, see Martin (2010, 2014) for a discussion regarding the historical classification and limitations of equities as a real asset. These investment characteristics are summarized in Exhibit 20.11 and discussed later relative to farmland and timber.

20.7.2 Risk and Return Expectations for Agriculture

Long-run return expectations for investments in core exposures (assuming little or no operating risk) in farmland, timber, and agriculturally oriented infrastructure tend to be modest, along with a similarly modest general risk profile of core investments in these asset classes. There may be substantial periods of outperformance (or underperformance) of these assets as changes in the macroeconomic environment favor or hinder particular investments in this area. Expected returns, as well as risks, may increase as investors assume greater operating exposure in these asset areas. There is not necessarily a monotonic relationship between expected return and stand-alone risk in real assets, as certain asset classes, such as commodities, may have low expected returns and higher risk, but may have additional characteristics that make them a good fit for an institutional portfolio:

- Ease of investment
- High liquidity
- Valuable role as an inflation hedge
- Positive diversification benefits
- Low interest rate sensitivity

Farmland, timber, and agriculturally related infrastructure benefit from:

- Relatively significant portion of their asset value made up of tangible value
- Relatively low operating risk required to utilize such assets
- Essential and substantial role they play in global economies

EXHIBIT 20.11 Comparison on Key Characteristics of Real Assets for Core Exposure in Various Asset Classes

Asset Class	Arguments for Being a Real Asset	Return Expectations	Risk	Ease of Investment	Liquidity	Inflation Hedge	Portfolio Diversifier	Interest Rate Sensitivity
Equities	Represent claims to real assets; nominal cash flows should adjust to changes in inflation	Medium	High	High	High	No	No	Medium
TIPS	Explicit linkage to CPI	Low	Low	High	High	Yes	No	High
Commodities	Core production inputs and scarcity	Low	High	High	High	Yes	Yes	Low
Real Estate	Essential component of economic infrastructure	Medium	Medium	Medium	Low	Uncertain	Yes	Medium-High
Timber	Scarce production input and biological growth	Low	Low	Medium	Low	Yes	Yes	Medium
Infrastructure	Essential component of economic infrastructure	Low	Low	Medium	Low	Uncertain	Yes	Medium
Gold	Store of value	Low	Medium	High	High	Yes	Yes	Low
Farmland	Essential component of economic infrastructure	Low	Medium	Medium	Medium	Yes	Yes	Medium-High
Intellectual Property	Essential component of economic infrastructure	Medium	Medium	Low	Medium	Uncertain	Yes	Low

Sources: Martin (2010); WCCM estimates.

20.7.3 Ease of Investment and Liquidity in Agriculture

Farmland, timberland, and agricultural infrastructure are typically accessed through managers who deploy capital through private market transactions, rather than the purchase of publicly listed securities. (There has been development of agriculturally related funds that make investments in listed equities, but this is not a common focus of institutional investors, who may already have exposure to this sector as part of broader equity-based allocations.) As private market transactions dominate these investments, ease of investment and liquidity are substantially lower than in listed securities. Actual liquidity certainly differs across the non–publicly traded real asset universe (e.g., row crop farmland located in active farming areas, which can readily be sold via regular and recurring auctions). For example, Sherrick (2012) estimates that approximately 2% of Illinois farmland turns over each year, though this is dominated by owner-operators. While no formal estimates have been made, the dollar value represented by this turnover is likely to be large relative to institutional investment in this area.

20.7.4 Inflation Hedge, Portfolio Diversification, and Interest Rate Sensitivity of Agriculture Investment

Farmland and timberland investment returns have well-documented positive correlation with U.S. CPI, and therefore can play a role in reducing the overall sensitivity of an investment portfolio to changes in inflation rates. In general, real assets represent portfolio diversifiers relative to portfolios dominated by stocks and bonds, as they typically have exposures less correlated with core beta. That said, many real assets involve interest rate sensitivity, either because they depend significantly on the cost of financing of ownership of real assets (e.g., real estate) through such mechanisms as lending, or they themselves generate cash flows (through rental, leasing, or other delegated usage, as well as from operating activity) over long horizons that themselves are capitalized into asset values at rates that depend in part on prevailing interest rates.

20.8 KEY POINTS

- Factor modeling of U.S. farmland shows at national (and state) levels that U.S. farmland has been positively correlated with inflation, and in that manner constitutes a real asset. Farmland has been negatively associated with the level of interest rates and positively associated with measures of economic growth.
- Factor modeling of U.S. farmland shows that core returns to farmland have been available with state-level investments.
- Agricultural infrastructure is an important part of the agricultural value chain, and can be an independent source of return.
- Substantial turnover in timberland assets to investment-oriented private owners has caused compression in timberland returns.
- Many institutional investors are looking toward less efficient non-U.S. timber and farmland assets, which offer the prospect of higher returns but also involve incremental ancillary risks, such as currency and legal risks.

- Farmland, timberland, and agriculturally related infrastructure have investment characteristics that can be different from those of other real assets. Such a diversity of investment characteristics is indicative of why institutional investors typically hold real assets from a variety of real asset classes.

20.9 CONCLUSION

Global demographic and economic trends support increased demand for agricultural products and correspondingly productive agricultural assets. While population and the demand for food will continue to grow, the supply of farmland is not likely to keep pace with this growth, and may even decline. Direct ownership of farmland has been an efficient means for the capital market investor to access returns to agriculture. An increasing number of institutional investors are turning to farmland and timberland to benefit from these trends as well as to increase the role of inflation protection in their portfolios.

NOTES

1. We acknowledge the coauthorship of George Martin. This chapter incorporates and adapts independent and original work (1) published by Geman and Martin (2011), which was sponsored by Bunge Global Agribusiness, whose support is gratefully acknowledged; (2) excerpted from Martin (2012); and (3) excerpted from materials that are part of an ongoing research program in agricultural infrastructure at Wood Creek Capital Management being led by George Martin, including WCCM (2015). Without suggesting any intellectual liability, George offers his warm thanks to his other coauthors: colleagues at Wood Creek Capital Management and the Center for International Securities and Derivatives Markets (CISDM) and the investment professionals at Bunge for supportive, intellectual, and practical engagement on farmland investing.
2. While well-known, the S&P GSCI may not be an efficient commodity index, particularly for accessing agricultural returns, and is therefore not fully representative of the agriculturally related returns available via futures markets. There may be other, more efficient indices available. Disclosure: Our coauthor has commercial relationships in this area.
3. This section incorporates previously published content from Martin (2010).
4. Estimated betas versus expected inflation range from –0.98 to +0.87, depending on the specific type of tree. An estimated beta of 1.0 for timber would indicate that an unlevered position in timber assets would serve as an effective inflation hedge only to the extent that the correlation coefficient between inflation and timber is high.
5. The study's authors abstract from land costs, since such costs are highly variable even within countries; instead, they focus exclusively on production and management costs.

REFERENCES

Ajanovic, A. 2011. "Biofuels versus Food Production: Does Biofuel Production Increase Food Prices?" *Energy* 36 (4): 2070–76.

Akers, K., and R. Staub. 2003. "Regional Investment Allocations in a Global Timber Market." *Journal of Alternative Investments* 5 (4): 73–87.

BP. 2014. "Statistical Review of World Energy."

Carpenter, J. E. 2010. "Peer-Reviewed Surveys Indicate Positive Impact of Commercialized GM Crops." *Nature Biotechnology* 28 (4): 319–21.

Cubbage, F., et al. 2010. "Global Timber Investments, Wood Costs, Regulation and Risk." *Biomass and Bioenergy* 24 (12): 1667–78.

Davis, H., et al. 2014. "Private Real Assets: Improving Portfolio Diversification with Uncorrelated Market Exposure." TIAA-CREF Asset Management.

Duvick, D. N. 2005. "The Contribution of Breeding to Yield Advances in Maize (*Zea mays* L.)." *Advances in Agronomy* 86:83–145.

FAO (Food and Agriculture Organization). 2013. "Emerging Investment Trends in Primary Agriculture."

———. 2015. FAOSTAT database.

Fu, C.-H. 2014. "Timberland Investments: A Primer." Timberland Investment Resources LLC. June. www.tirllc.com/wp-content/themes/tirllc/docs/TIR_A-Primer-2012-06-11-02.pdf.

Gardner, B. 2003. "U.S. Commodity Policies and Land Values." In *Government Policy and Farmland Markets*, edited by C. Moss and A. Schmitz, 81–96. Ames: Iowa State Press.

Geman, H., and G. Martin. 2011. "Understanding Farmland as Part of a Diversified Portfolio." Bunge Asset Management.

Goodwin, B., et al. 2003. "Explaining Regional Differences in the Capitalization of Policy Benefits into Agricultural Land Values." In *Government Policy and Farmland Markets*, edited by C. Moss and A. Schmitz, 97–114. Ames: Iowa State Press.

Hay, R., and J. Porter. 2006. *The Physiology of Crop Yield*. 2nd ed. Oxford, UK: Blackwell Publishing.

Healy, T., et al. 2005. "Timber as an Institutional Investment." *Journal of Alternative Investments* 8 (3): 60–74.

IEA (International Energy Agency). 2014. "IEA World Energy Outlook." www.worldenergyoutlook.org.

Ifft, J., T. Kuethe, and M. Morehart. 2015. "The Impact of Decoupled Payments on U.S. Cropland Values." *Agricultural Economics*.

Kastens, T. 2001. "*Risk and Reward: How Do Farm Returns Stack Up? Should Farm Managers Invest in the Stock Markets?*" Kansas State University.

Kay, S., J. Peuch, and J. Franco. 2015. "Extent of Farmland Grabbing in the EU." European Parliament Committee on Agriculture and Rural Development.

Kim, D. 2012. "Is Currency Hedging Necessary for Emerging Market Investments?" *Economics Letters* 116 (1): 67–71.

Kirwan, B., and M. Roberts. 2010. "Who Really Benefits from Agricultural Subsidies: Evidence from Field Level Data." University of Maryland.

Klümper, W., and M. Qaim. 2014. "A Meta-Analysis of the Impacts of Genetically Modified Crops." *PLOS One* 9 (11): e111629.

Linehan, V., S. Thorpe, N. Andrews, Y. Kim, and F. Beaini. 2012. "Food Demand to 2050: Opportunities for Australian Agriculture." Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) Conference Paper 12.14. March.

Lonnstedt, L., and R. Sedjo. 2012. "Forest Ownership Changes in the United States and Switzerland." *Forest Policy and Economics* 14:19–27.

Martin, G. 2010. "The Long-Horizon Benefits of Traditional and New Real Assets in the Institutional Portfolio." *Journal of Alternative Investments* 13 (1): 6–29.

———. 2012. "*Commodity Price Volatility and Farmland-Based Investment Strategies*." Presentation. University of London, Birkbeck, January 2014.

———. 2014. "Real Assets: An Institutional Perspective." Presentation. Center for International Securities and Derivatives Markets (CISDM). October 2014.

Meketa Investment Group. 2010. "Timberland." April. www.meketagroup.com/documents/TimberlandWP_000.pdf.

Morgan Stanley. 2013. "Mid-Stream Energy MLP's Primer 3.0." April 17.

NBIM (Norges Bank Investment Management). 2013. "Infrastructure Investments." NBIM Discussion Note 2-2013.

Peters, C. J., J. L. Wilkins, and G. W. Fick. 2007. "Testing a Complete-Diet Model for Estimating the Land Resource Requirements of Food Consumption and Agricultural Carrying Capacity: The New York State Example." *Renewable Agriculture and Food Systems* 22 (02): 145–53.

R&A Investment Forestry. 2014. "U.S. Timberland for the Private Investor."

Sherrick, B. 2012. "Farmland Turnover in Illinois." University of Illinois–Urbana-Champaign farmdoc Project. November 29. http://farmdocdaily.illinois.edu/2012/11/farmland-turnover-in-illinois.html.

Smith, S., et al. 2014. "Yield Gains in Major U.S. Field Crops." CSSA Special Publication 33:125–71.

Sohn, I. 2007. "Long-Term Energy Projections: What Lessons Have We Learned?" *Energy Policy* 35 (9): 4574–84.

Specht, J., et al. 2014. "Soybeans." In Smith et al. (2014).

Suhre, J., et al. 2014. "Soybean Yield Partitioning Changes Revealed by Genetic Gain and Seeding Rate Interactions." *Agronomy Journal* 106 (5): 1631.

United Nations Population Division. 2008. "World Population Prospects: The 2008 Revision Population Database." http://esa.un.org/unpp/index.asp.

Vegetarian Society. 2009. "Why It's Green to Go Vegetarian." www.vegsoc.org/environment/land_use.html.

Waggle, D., and D. T. Johnson. 2009. "An Analysis of the Impact of Timberland, Farmland and Commercial Real Estate in the Asset Allocation Decisions of Institutional Investors." *Review of Financial Economics* 18 (2): 90–96.

Washburn, C., and C. Binkley. 1993. "Do Forest Assets Hedge Inflation?" *Land Economics* 69 (3): 215–24.

Washburn, C., P. D'Anieri, and M. E. Aronow. 2004. "What Do Public Markets Tell Us about Timberland Property Values?" *Hancock Timberland Investor*, Third Quarter.

WCCM (Wood Creek Capital Management). 2015. "Agricultural Infrastructure for Institutional Investors: Productivity, Geography and Economic Returns."

Wisner, R. 2015. "Estimated U.S. Dried Distillers Grains with Solubles (DDGS) Production and Use." www.extension.iastate.edu/agdm/crops/outlook/dgsbalancesheet.pdf. Accessed April 10, 2015.

CHAPTER 21

Investing in Intellectual Property

A substantial portion of gross domestic product (GDP) is now generated by and composed of so-called intangible assets, such as intellectual property (IP).[1] Like tangible assets, these assets are important ingredients in the production of goods, and, as previously discussed, certain IP assets have many of the same characteristics as some traditional real assets. This chapter provides an overview of IP and breaks it down into three categories: (1) film production and distribution, (2) visual works of art, and (3) patents and research and development (R&D).

21.1 CHARACTERISTICS OF INTELLECTUAL PROPERTY

Historically, most intangible assets were bundled with other corporate assets and available for investment through traditional means, such as an equity investment in a software (or other technology-oriented) company. However, in recent years, there has been an increased interest in unbundling and isolating intangible assets, IP in particular, for stand-alone investment purposes. Examples of such assets include patent portfolios, film copyrights, art, music or other media, research and development (R&D), and brands.

Unbundled intellectual property is IP that may be owned or traded on a stand-alone basis. Unbundled IP may be acquired or financed at various stages in its development and exploitation. Ex ante, newly created IP may have widely varying value and use. The value of property such as exploratory research, new film production, new music production, or pending patents will typically be widely uncertain prior to production or implementation. Similar to venture capital investments, many of these types of IP may fail to recapture initial investment or costs, while a proportionately small number of cases will capture a large asymmetric return on investment. For example, De Vany and Walls (2004) report the received wisdom that "80% of movies lose money"; yet films, on average, are still profitable, and in the subsample of films studied, far more than 20% are profitable. De Vany and Walls (2004) report that, for a sample of more than 2,000 films, 6.3% of the films generated 80% of the total profits.

Mature intellectual property is IP that has developed and established a reliable usefulness. Mature IP will typically have a much more certain valuation and a more clear ability to generate licensing, royalty, or other income associated with its use. Such income, to the extent that it has returns and risks that are understood, is likely to be well priced in the marketplace, with risk premiums associated with asset complexity, asset liquidity, and variability of asset returns.

Returning to the example of film, Soloveichik (2010) argues that a theatrical film should be considered a long-lived asset. Life spans are typically 80 years, during which 50% of a film's value is lost in the first year after release, and depreciation in subsequent years is about 5% per year of beginning period value. In addition to predictable depreciation, post–box office revenues tend to be correlated with box office revenues. While it is likely somewhat a function of statistical outliers, it's worth noting that De Vany and Walls (2004) find, perhaps implausibly,[2] that domestic DVD sales and licensing revenues for their sample of films have a correlation of 0.97 and 0.99, respectively, with domestic box office revenues.

To the extent that specific IP is primarily a consumption good and largely removed from the capital accumulation process (as is typically the case with pure works of art or other noncommercialized media), economic reasoning and empirical evidence suggest that such assets as a whole are less likely to generate substantial and sustainable long-term investment returns.[3] However, the evidence (Nakamura 2009) does suggest that there is substantial capital accumulation for many intangible assets, which implies that intangible assets can represent important investment opportunities.[4]

Evidence also exists that IP may offer substantial investment returns. For example, estimates of the private, pre-tax returns to R&D expenditures are in the 20% to 30% range (although it is important to note that there are substantial issues in measuring these returns). Furthermore, inclusion of the spillover effects elevates the estimated rate of return significantly.[5] The **spillover effects** of an activity, also known as externalities, represent effects on other entities, such as benefits realized by other firms and consumers from the successful R&D of a firm. The National Science Foundation estimates that total corporate domestic R&D in 2013 was $322 billion, with 69% of R&D being conducted by the manufacturing sector. Pharmaceuticals represent about 16% of total R&D expenditures, computers and electronic products represent another 21%, and software represents 11%.[6]

The next three sections discuss the major categories of IP for which there are well-developed bodies of economic research: (1) film production and distribution, (2) visual works of art, and (3) R&D and patents.

21.2 FILM PRODUCTION AND DISTRIBUTION

Film production and distribution comprise a subset of IP that often has relatively substantial accounting data availability and thus provides a good example of the methodologies for estimating and modeling expected future cash flows and accounting profitability.

21.2.1 Film Production and Distribution Revenues

Film production and distribution fall into the IP category of artwork. Soloveichik and Wasshausen (2011) report that in 2007, the total revenue from the sale and licensing of copyrighted artwork in the United States totaled $285 billion. Total revenues from film production (including exports of U.S.-produced films and U.S. revenues

EXHIBIT 21.1 Schedule of Film Exhibition Venues

Exhibition Form	Window	Time after Release
Theatrical	6 months	0
Home Video	10 years +	4 months
Pay-per-View	2 months	8 months
Pay TV	18 months	12 months
Network	30 months	30 months
Pay TV Second Window	12 months	60 months
Basic Cable	60 months	72 months
Television Syndication	60 months	132 months

Source: Wood Creek Capital Management (2011).

generated from non-U.S. films), including exhibition, licensing for home media and broadcast, and ancillary income, were estimated to be $37 billion.[7] Film revenues are generated almost exclusively by exhibition, which has a generally stable set and sequence of stages, though not all films will be licensed for exhibition in all forms. Exhibit 21.1 shows the schedule of exhibition venues.

While total revenues from film have demonstrated relative stability, the mix of revenue sources has been changing relatively quickly, due in part to technology but also to other financial imperatives, such as the availability (or lack) of capital for new film production. Examples of changing revenue sources include the rise and subsequent relative decline in revenues associated with DVD and similar exhibition technologies, as well as the increasing importance of online and non-U.S. revenues to overall revenue.

21.2.2 Film Production and Distribution Life Cycle

In order to understand the opportunities for generating returns from film production and distribution, especially in a relatively dynamic period, it is helpful to review the life cycle of a film. **Film production stages** are:

- Story rights acquisition: Payments to license concepts, books, or screenplays
- Preproduction: Script development, set design, casting, crew selection, costume design, location scouting, and budget
- Principal photography/production: Compensation of actors, producers, directors, writers, soundstage, wardrobe, set construction, labor, catering, and lodging
- Postproduction: Film editing, scoring, titles and credits, dubbing, special effects, and soundtrack music rights or composition

21.2.3 Costs and Financing of Film Production and Distribution

Collectively, the costs of film production are often referred to as negative costs, which refer not to the sign of values but to the fact that these are all costs required to produce what was, in the predigital era, the film negative. These costs are coupled with the

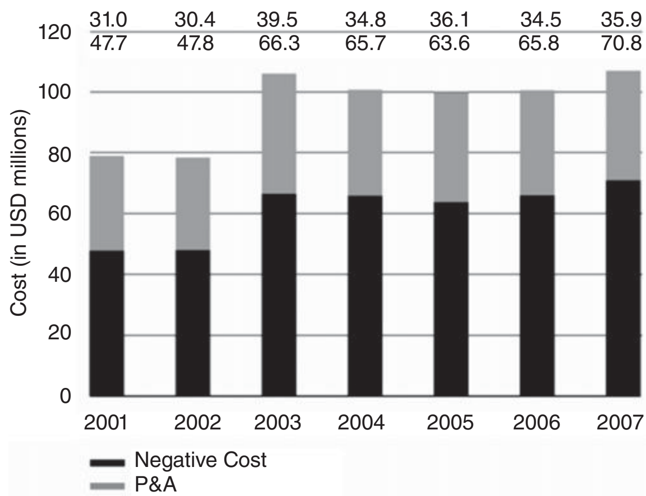


EXHIBIT 21.2 MPAA Member Production and Marketing Costs (per Film)
Source: Motion Picture Association of America (MPAA).

substantial cost of prints and advertising (P&A), which is the cost of the film prints to be used in theaters (whether digital or physical) and the film's advertising and marketing costs (see Exhibit 21.2).[8]

Revenues from each of these exhibition forms are used to pay for the financing required to produce and distribute a film. Financing is achieved through equity or debt financing, or a combination of both.

Equity financing structures include slate equity financing, corporate equity, coproduction, and miscellaneous third-party equity financing:

- Slate equity financing: In **slate equity financing**, an outside investor (e.g., hedge fund or investment bank) funds a set of films to be produced by a studio. These slates typically reflect a set of parameters regarding diversification, risk, the number of films to be released, minimum and maximum budgets for film production and P&A, and genre diversification requirements. Slate deals emerged to spread financial risk across a series of films, thus limiting the impact of one film's losses on an overall financial investment. Slate financings may have further provisions to ensure against moral hazards; such provisions would, for example, deter a studio from assigning films with lower expected returns or greater risk to slate financings.
- Corporate equity: This is equity fund-raising (private placement or public offering) to fund the activities of a production company.
- Coproduction: In **coproduction**, two or more studios partner on a film, sharing the equity costs and, correspondingly, the risks and returns.
- Miscellaneous third-party equity: Some combination of high-net-worth individuals, institutional investors, and other third-party investors fund costs not covered by other types of financing; this is particularly common for smaller independent films.

Debt financing structures include senior secured debt, gap financing, and super gap financing/junior debt:

- Senior secured debt: A bank or another financial institution lends funds to a movie studio or producer to finance the production and/or P&A of a film. This loan can come in various structures and forms, backed by specific collateral, such as the following:
  - *Negative pickup deal.* A **negative pickup deal** occurs when a film distributor agrees to purchase a film from a producer for a fixed sum upon delivery of the completed film.
  - *Foreign presales*. A **foreign presale** occurs before the film is made, when the producer sells distribution rights for specific foreign territories for a fixed price; all, or nearly all, of this payment is due upon delivery of the completed film.
  - *Tax credits/grants*. The producer receives tax credits (which are salable) or grants (paid in cash) for filming in a specific state or country.
- Gap financing: **Gap financing** covers the difference between the production budget and the senior secured debt, which can be collateralized by sales of unsold territories to distributors.
- Super gap financing/junior debt: **Super gap financing** is a second level of gap financing, often syndicated, representing the final gap that the senior lender or gap financier does not want to risk.

Further, financing may be supported directly or indirectly with royalty participations. These may, in the case of talent, be in lieu of salary or other noncontingent compensation; or, in the case of financial investors, they may be used to lower the cost of up-front financing. These participations are usually assignable (i.e., transferable to third parties) after a film has been produced.

21.2.4 Four Findings from Evidence on Film Production Profitability

Economists have long tried to bring some order to the study of ex ante determination of box office revenues (a world in which, as screenwriter William Goldman famously noted, "nobody knows anything"). The typical focus of this analysis has been on domestic box office revenue (DBOR), since this information is readily available in near real time.[9] More recently, international box office revenue (IBOR) data have become generally available; however, detailed information on post-theatrical revenues is typically unavailable to the public on a film-by-film basis, although particular stylized facts are discussed or deduced from aggregates, such as the fact that post-theatrical revenues are typically similar to box office revenues.

Translating revenue numbers into profits is generally impossible without direct knowledge of and participation in the production of particular film assets. However, there are many regularities that arise in contracting, and these can be used to conduct analysis. In particular, it is a relatively conventional (and accurate) assumption that theatrical exhibiters earn roughly 50% of the take at the box office, and this proportion of the take will tend to increase with the length of time that a film is on the exhibitor's display schedule.

A growing literature has looked at the impact of various factors in determining revenue or profit associated with exhibition, with various studies including different stages of the exhibition life cycle as well as more or less actual costs incurred for big items like P&A. There are four key conclusions from the empirical evidence, summarized in Exhibit 21.3:

1. Bigger budgets tend to lead to bigger revenues but lower average profitability.
2. Some film stars may have a positive impact on movie revenues, but most do not. John, Ravid, and Sunder (2014) provide evidence that directors may have an impact on film profitability.
3. Sequels tend to generate greater revenues and lower risk.
4. Different genres have different risk-return properties.

Relatedly, there is some question as to whether the returns to a film can be well determined based on opening box office revenues. De Vany and Walls (1999) find they cannot and provide evidence that there is a critical inflection point, at around six weeks after box office opening, when film attendance either trails off or expands significantly.

21.2.5 Estimating the Relationship of Returns to Film Production

A simple metric of the return to an investment is the ratio of the cash in to the cash out. A ratio of –1.0 indicates no revenues, an investment with a ratio of 0.0 breaks even (ignoring the time value of money), and a positive ratio indicates a profit. This section examines the historical returns (the "cash in to cash out" ratios) to various film categories and fits curves to those returns in an attempt to provide analysts with simplified summaries of their behavior. The challenge is to identify a type of distribution that is able to closely approximate the observed frequencies of various levels of profitability.

For example, profitability of films is widely recognized to be highly skewed to the right, similar to venture capital (see Exhibit 21.4). In other words, there are many outcomes with little or no profitability, and a small percentage of films with very high profitability. Obviously the normal distribution, which is symmetrical, would serve as a very poor distribution with which to model film investments. Previous academic work has attempted to model the wildly uncertain character of box-office-related revenues using the fat-tailed stable Paretian distribution, which is often referred to as the Pareto principle, or the 80/20 rule.[10] However, the cash flows for box office revenues have been observed to be far more skewed than can be modeled by the stable Paretian distribution.[11]

The four-parameter kappa distribution (K4) of Hosking (1994) is commonly used to model distributions with fat tails. Of particular usefulness is the ability of the K4 distribution to assume a wide variety of shapes with fat tails, including densities with no bounds, lower bounds, or lower and upper bounds. The K4 distribution is parameterized with two shape parameters, one scale parameter, and one location parameter. This section demonstrates the use of the K4 distribution to model the returns to films. Data are collected from the 981-film Opus Data set for the period 2005–10, including DBOR, global box office revenue (GBOR), production costs, film

EXHIBIT 21.3 Academic Literature on Factors Determining Revenues/Profits to Film Exhibition

Factor	Study	Dependent Variable	Findings
Budgets	Litman (1983), Litman and Kohl (1989), Litman and Ahn (1998), Zufryden (2000), Elberse and Eliashberg (2003)	Revenues	Budgets are positively associated with revenues.
	John, Ravid, and Sunder (2002), Hennig-Thurau, Houston, and Walsh (2007)	Profits	Movies with larger budgets are less profitable.
	Ravid (2004), Ravid (1999)		Large budgets may decrease profitability.
Stars	Liu (2006), De Vany and Walls (1999), Ravid (1999), Litman (1983)	Revenues	There is no relationship between star power and revenues.
	Basuroy, Desai, and Talukdar (2006), Ainslie, Dreze, and Zufryden (2005), Elberse and Eliashberg (2003), Basuroy, Chatterjee, and Ravid (2003), Neelamegham and Chintagunta (1999), Sawhney and Eliashberg (1996), Sochay (1994), Litman and Kohl (1989)		There is a positive relationship between star power and opening or total revenues.
	De Vany and Walls (1999), DeVany and Walls (2002), Ravid (1999), Ravid and Basuroy (2004)	Profits	There is no relationship between star power and profits.
	Hennig-Thurau, Houston, and Walsh (2007)		There is a slightly negative association between star power and profits.
Production-studio-related effects	Litman and Kohl (1989)	Revenues	Top directors increase revenues.
	Ainslie, Dreze, and Zufryden (2005)		Revenues and market shares vary across studios, due to different levels of expertise in timing of releases and accuracy about predicted competition.
	Hennig-Thurau, Houston, and Walsh (2007)	Profits	Director power does not guarantee increased profits.
Sequels	Ravid (1999)	Revenues	Sequels are associated with higher revenues.
	Ainslie, Dreze, and Zufryden (2005)		Sequels are associated with stronger openings.
	Basuroy and Chatterjee (2008)		Sequels earn less revenue than originals, but more than contemporaneous nonsequel movies.
	Hennig-Thurau, Houston, and Heitjans (2009)		Sequels are associated with higher revenue and less risk.
	Ravid (1999)	Profits	Sequels are associated with marginally higher profitability.
	Walls (2006)		Sequels have higher expected profitability.
Genres	Elberse and Eliashberg (2003)	Revenues	Literature review concludes that findings with respect to genre are inconclusive.
	Ainslie, Dreze, and Zufryden (2005)		Very few genre variables are significant.
	Litman (1983)		Science fiction genre is associated with higher revenue.
	Neelamegham and Chinatagunta (1999)		Thriller (romance) is associated with higher (lower) revenue.

Source: Adapted from Karniouchina, Carson, and Moore (2010).

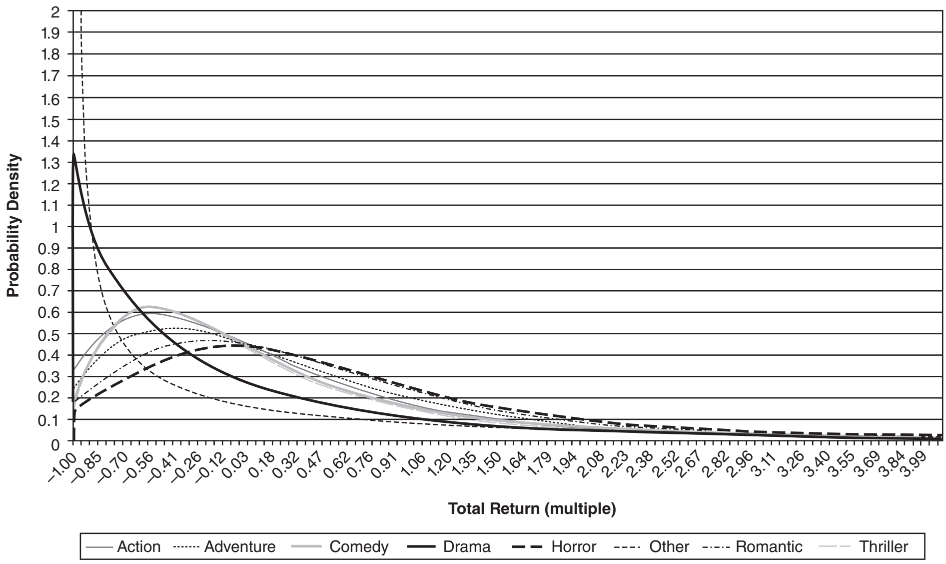


EXHIBIT 21.4 Probability Density of Estimated Total Box Office Returns by Genre (K4 model), 2005–10

genre, and date of release. P&A spending is not considered on the cost side, and post-theatrical revenues are not considered on the revenue side:

$$\text{Estimated Cash-on-Cash Returns} = \{[(\text{DBOR} + \text{GBOR})/2]/\text{Production Costs}\} - 1$$

The probability densities of returns to various film genres are graphed as shown in Exhibit 21.4 using the K4 distribution. In other words, the four parameters of the K4 distribution are econometrically estimated and then used to form the graphs that approximate the relationship between probability and return for various film genres.

From the density functions, action, drama, and “other” have the most skewness. The estimated density functions can also be used to estimate the probability that each genre will fail to break even (i.e., generate a loss). Exhibit 21.5 provides the estimated probabilities of loss for each film genre.

Exhibit 21.5 indicates a range of probabilities for different genres, with horror the least risky and “other” the most risky. While still displaying substantial risk, the results are much less pessimistic than suggested by the received wisdom.[12] This is partly a result of the growth in outlets for film production, particularly online and in international markets. It is also likely the result of better segregation of studio overhead from direct production costs, as well as increasing efficiency in the production process, which is associated, in part, with competitive pressures between studios. More generally, because of the asymmetry and fat tails associated with box office returns, traditional investment performance and risk measures (e.g., correlations, Sharpe ratios, regression betas) are generally invalid, and the analyst interested in constructing and monitoring portfolios of such risks must address this directly.

Pitt (2010) offers a book-length treatment of issues relating not just to music sales but also to royalty revenue streams associated with music production and publishing.[13] He finds a similar dynamic in music royalty streams as exists in

EXHIBIT 21.5 Estimated Probability That Films Do Not Break Even

Genre	Prob X < 0
Action	0.56
Adventure	0.49
Comedy	0.53
Drama	0.66
Horror	0.37
Musical	N/A
Other	0.72
Romantic	0.42
Thriller	0.55

Source: Authors' calculations.

movies—a small number of performers, songwriters, and copyright owners generate the bulk of total revenues—and that the statistical distribution of revenue per work is significantly non-normal and possesses fat tails. Pitt also finds that other factors, such as the total number of songs in an artist's catalog and longevity, can have an impact on total royalty revenues per artist.

21.3 VISUAL WORKS OF ART

Over the past 20 years, there has been a growing interest in understanding the investment properties of art, due in part to investors' increasing interest in considering art acquisition and holding as part of a broader wealth-management strategy. Within IP, art tends to provide superior pricing data for analysis of historical returns.

21.3.1 Reasons for Considering Art as an Investment

Capgemini (2011) reports that 42% of wealth advisers to high-net-worth families believe that their clients purchase art primarily as a strategy for accumulating wealth. Capgemini and RBC Wealth Management (2013) report that 17% of investor allocation to "investments of passion" is to art (with jewelry, gems, and watches being the largest category, at 32%). Moreover, there is much practitioner literature emphasizing that the newly wealthy have a demand for art that is driving up auction prices for art, particularly for certain styles and traditions.

More generally, there are a number of reasons advanced for considering art as an investment:

- Long, sustained history of substantial demand by investors and consumers
- Elevated potential for market inefficiency and therefore mispricing, due to illiquidity and private valuation
- Lack of transparency in auction markets[14]

However, the question remains: Does the empirical evidence indicate that holding broad baskets of art earns compelling long-run returns? In other words, to what extent are investors, perhaps subconsciously, involved in the asset class for entertainment value related to their passion for art versus the desire to profit financially?

There are a number of immediate issues that arise when trying to understand the investment performance of art as an asset class. In particular, unlike traditional financial assets, which have continuous markets and relatively homogeneous sources of risk, art markets are characterized by infrequent transactions in highly heterogeneous assets. Without some special techniques, merely looking at weighted averages of infrequently traded art prices will induce substantial smoothing in returns. Similarly, disproportionate turnover in one subset of art will bias any inference about the prices of the overall universe of art being considered. As a practical matter, index construction methods to deal with these issues can be divided into two groups: (1) hedonic price estimators, which use a regression-based methodology to synthetically develop continuous price series by controlling for the unique characteristics (qualities) of each transaction that comes to market, and (2) repeat-sales estimators, which use a regression-based methodology that focuses on the returns to works of art that have more than one transaction. Both of these methods, and their extensions, have their origins in the efforts of econometricians to understand other illiquid and heterogeneous assets, such as real estate.[15]

21.3.2 Evidence on the Investment Returns to Art

There is a wide range of studies regarding the returns to art, typically by style or geography. For example, Taylor and Coleman (2011) create price and return indices for Australian Aboriginal art. Ashenfelter and Graddy (2003) offer an extensive review of the literature through the early 2000s, which is broadly representative of subsequent research on returns. The results of the review are summarized in Exhibit 21.6.

It can be seen in Exhibit 21.6 that the median real return to holding art over extended periods of time is 2.2% (return estimates do not vary substantially by estimation methodology). However, most studies of the returns to art investment consider only **hammer prices**, which are final auction prices that do not include commissions to the auction house. Commissions that may be charged to both the buyer and the seller amount to as much as 15%. If we assume that the typical round-turn transaction cost for a sale is 25%, then it would be expected to take 10 years of price appreciation to cover the transaction costs associated with a piece of art.

21.3.3 Characteristics Hypothesized to Drive Returns to Art

Ashenfelter and Graddy (2006) offer a further review of the literature designed to evaluate the presence of two other possible characteristics of the art market that could conceivably form the basis for investment strategies. The first, known as the **masterpiece effect**, hypothesizes that the returns to the most expensive artworks are qualitatively different from the market as a whole. Ex ante, one should consider if such an effect should exist and, if so, what sign it should have. Of six studies reviewed by Ashenfelter and Graddy (2006), only one finds a positive masterpiece effect. The second market characteristic in question is the so-called law of one price: Do transactions in similar works of art across auction houses and/or geographies occur within the bounds set by transaction costs? Of the various studies reviewed, there is

EXHIBIT 21.6 Estimated Returns to Art from Various Studies

Author	Sample	Period	Method	Nominal Return	Real Return
Anderson (1974)	Paintings in general	1780–1960	Hedonic	3.3%	2.6%
	Paintings in general	1780–1970	Repeat sales	3.7%	3.0%
Stein (1977)	Paintings in general	1946–1968	Assumes random sampling	10.5%	
Baumol (1986)	Paintings in general	1652–1961	Repeat sales		0.6%
Frey and Pommenihne (1989)	Paintings in general	1635–1949	Repeat sales		1.4%
		1950–1987	Repeat sales		1.7%
Buelens and Ginsburgh (1993)	Paintings in general	1700–1961	Hedonic		0.9%
Pesando (1993)	Modern prints	1977–1991	Repeat sales		1.5%
Goetzmann (1993)	Paintings in general	1716–1986	Repeat sales	3.2%	2.0%
Barre et al. (1996)	Great impressionist	1962–1991	Hedonic	12.0%	5.0%[a]
	Other impressionist	1962–1991	Hedonic	8.0%	1.0%[a]
Chanel et al. (1996)	Paintings in general	1855–1969	Hedonic		4.9%
	Paintings in general	1855–1969	Repeat sales		5.0%
Goetzmann (1996)	Paintings in general	1907–1977	Repeat sales		5.0%
Pesando and Shum (1996)	Picasso prints	1977–1993	Repeat sales	12.0%	1.4%
Czujack (1997)	Picasso printings	1966–1994	Hedonic		8.3%
Mei and Moses (2001)	American, Impressionist, and old master	1875–2000	Repeat sales		4.9%
Graeser (1993)	Antique furniture	1967–1986	Neither[b]	7.0%	2.2%
Ross and Zondervan (1989)	Stradivarius violins	1803–1986	Hedonic		2.2%

[a]As many of the surveys report only nominal returns, the authors calculated the real return rates as follows. For the Anderson and Baumol studies, an inflation rate of 0.7% a year was used. This number is based on Baumol's estimate of inflation during the 300-year period of his study using the Phelps-Brown and Hopkins price index. Goetzmann's estimate of inflation during the period of his study (also based on Phelps-Brown and Hopkins) is 1.2%. French price inflation between 1962 and 1992 according to OECD statistics was 7%.
[b]Assumes random sampling within a portfolio of fixed furniture types.
Source: Ashenfelter and Graddy (2003).

EXHIBIT 21.7 Real Returns to Art

Country	Period	Annual Return	Volume	Minimum	Maximum	Annual GDP	Equity Return
Australia	1971–2007	3.09%	21.15%	–40.40%	66.66%	3.91%	7.56%
Austria	1971–2007	2.53%	17.44%	–38.40%	37.76%	4.11%	5.66%
Belgium	1975–2007	–0.90%	17.41%	–44.30%	38.57%	2.02%	8.42%
Canada	1972–2007	2.36%	16.12%	–27.19%	28.31%	3.03%	5.91%
Denmark	1976–2007	1.75%	15.56%	–41.96%	29.66%	2.70%	9.58%
France	1971–2007	1.14%	18.94%	–61.03%	39.74%	3.50%	8.00%
Germany	1971–2007	1.52%	13.12%	–28.08%	25.87%	1.43%	5.44%
Italy	1971–2007	1.99%	17.67%	–41.80%	36.66%	3.68%	6.75%
Netherlands	1971–2007	2.30%	17.94%	–44.26%	48.74%	4.13%	8.14%
Sweden	1971–2007	2.32%	20.18%	–55.16%	44.56%	2.65%	10.51%
Switzerland	1972–2007	1.99%	18.50%	–38.99%	75.63%	3.41%	7.30%
UK	1971–2007	4.60%	15.79%	–38.66%	30.47%	3.88%	7.94%
U.S.	1971–2007	3.07%	14.31%	–28.73%	33.58%	2.53%	6.19%

Note: Results are USD denominated, deflated.
Source: Spaenjers (2010).

weak evidence that the law of one price does not hold—for example, for transaction prices on the works of Picasso, as they vary systematically between auction locations (London versus Paris).

In a series of related papers, Spaenjers and coauthors offer a current, comprehensive view of the international art market. In particular, Spaenjers (2010) considers data on over one million art transactions across 13 countries from the 1960s onward, accounting for both geographical and currency effects. The results, summarized in Exhibit 21.7, are consistent with the preceding: Annualized real returns to a diversified basket of art have been in the neighborhood of 2% and do not vary significantly across geographies or markets. In addition, the volatility of art indices has a median of 17% per year. This combination of risk and return compares unfavorably to historical experience in equity markets.

One immediate question that arises is the extent to which these returns are a function of fluctuations in the value of the U.S. dollar. Spaenjers (2010) provides domestic currency returns for the same 13 countries, which have a median return of 1% and a correlation of 0.88 with the annualized returns reported previously. This suggests that currency effects are not a significant driver of art market returns. Spaenjers (2010) also looks at quality effects, in which works of better-known artists are considered to be of higher quality. From the adjacent results, a mild **quality effect** is noted, in which higher-quality paintings generate higher returns. However, these higher returns were at the expense of higher volatility. Median information ratios for highest-quality paintings are 0.23, with corresponding information ratios of 0.11 and 0.08 for medium- and lower-quality paintings. From an investment perspective, these information ratio–based findings are unattractive.

With such low returns, why might it be that there is significant demand for art? Spaenjers (2010) offers preliminary but intuitive evidence, summarized in Exhibit 21.8, that art prices are significantly explained by wealth effects, as proxied by gross domestic product (GDP) growth, lagged equity market effects, and income

EXHIBIT 21.8 Risk and Return for Art Investments, Sorted by Quality

		High Quality		Medium Quality		Low Quality	
Country	Period	Return	Volatility	Return	Volatility	Return	Volatility
Australia	1980–2007			4.23%	23.13%	3.98%	17.97%
Austria	1980–2007			2.37%	20.80%	0.32%	17.25%
Belgium	1980–2007			0.60%	20.80%	–0.64%	16.57%
Canada	1980–2007					2.13%	17.34%
Denmark	1980–2007			1.32%	20.27%	0.90%	16.26%
France	1980–2007	3.21%	28.94%	1.39%	19.34%	1.57%	19.01%
Germany	1980–2007			0.96%	13.71%	–0.36%	14.55%
Italy	1980–2007			2.93%	17.04%	2.83%	18.21%
Netherlands	1980–2007			2.57%	18.81%	1.28%	14.60%
Sweden	1980–2007			1.56%	25.33%	–0.05%	19.93%
Switzerland	1980–2007					–0.72%	14.82%
UK	1980–2007	5.95%	17.62%	4.75%	15.23%	4.03%	14.06%
U.S.	1980–2007	4.85%	20.95%	3.78%	14.49%	3.25%	11.13%

Note: Results are USD denominated, deflated.
Source: Spaenjers (2010).

inequality. While interesting in their own right, these results also suggest that the diversification benefit of art is lower than some may anecdotally believe, as most traditional investors already have positive exposure to GDP and equity markets; high-net-worth investors will also have existing positive exposure to wealth inequality.

Forsyth (2012) suggests that high-net-worth investors invest in art as a hedge against inflation or confiscation of wealth by governments. For those with a net worth above $100 million, he suggests, an important goal is to maintain rather than grow wealth. Artworks can protect against monetary debasement, confiscation, and social unrest. Forsyth quotes Richard Morais: “Any private banker will tell you that, as soon as a centimillionaire … makes their fortune, the first thing they do is figure out how they can ferret away large chunks of that wealth to countries that guarantee political and personal freedoms, have sound legal systems, a favorable tax environment, good security and good schools for their kids.” A substantial portion of this newfound wealth may be invested in real estate in cities such as New York or London, and in art, which can be easily shipped to the residences in these safe, global cities.

Another explanation of low financial returns to art could be that the investment in art provides a total return that is a combination of the financial return to art (price appreciation) and the aesthetic benefit to being the owner of the art. The aesthetic (nonfinancial) benefit to owning art includes the joy of viewing and otherwise controlling the art. To the extent that competition drives the total return to similar risk-adjusted levels, there is a trade-off between the financial return and the aesthetic benefit. In artwork overall, and perhaps in some artwork in particular, prices are driven higher (and expected financial returns are driven lower) in anticipation of the nonfinancial benefits from ownership.

Are the relatively low monetary returns to art, particularly given risk levels, unique to art or more broadly a characteristic of collectibles, including stamps and rare violins? Dimson and Spaenjers (2014) provide documentary evidence that

EXHIBIT 21.9 Risk and Return of Art and Other Collectibles, 1900–2012

	Mean Annual Returns			Range of Annual Returns			
	Geometric	Arithmetic	St. Dev.	Lowest		Highest	
Nominal Returns							
Art	6.4%	7.2%	13.2%	–31.2%	1930	46.6%	1968
Stamps	6.9%	7.6%	13.5%	–8.8%	1982	83.2%	1979
Violins	6.5%	7.0%	10.1%	–22.4%	2011	41.0%	1974
Equities	9.4%	11.2%	21.6%	–48.8%	1974	145.6%	1975
Bonds	5.5%	6.1%	11.9%	–17.4%	1974	53.1%	1982
Bills	4.9%	5.0%	3.8%	0.3%	2012	17.2%	1980
Gold	5.1%	6.4%	18.7%	–19.0%	1990	108.3%	1979
Inflation	3.9%	4.2%	6.5%	–26.0%	1921	24.9%	1975
Real Returns							
Art	2.4%	3.1%	12.4%	–29.7%	1915	38.4%	1968
Stamps	2.8%	3.5%	12.3%	–19.2%	1915	56.3%	1979
Violins	2.5%	2.8%	8.5%	–25.9%	2011	23.9%	1986
Equities	5.2%	7.1%	19.8%	–57.1%	1974	96.7%	1975
Bonds	1.5%	2.4%	13.7%	–30.7%	1974	59.4%	1921
Bills	0.9%	1.1%	6.3%	–15.7%	1915	43.0%	1921
Gold	1.1%	2.2%	16.6%	–30.5%	1975	77.7%	1979

Correlations

	Art	Stamps	Violins	Equities	Bonds	Bills	Lagged Equities
Real Returns							
Art	—						0.34
Stamps	0.14	—					0.20
Violins	0.25	0.07	—				0.14
Equities	0.22	0.00	0.02	—			–0.07
Bonds	0.08	0.24	0.03	0.51	—		–0.11
Bills	0.23	0.36	0.35	0.26	0.63	—	0.06
Gold	0.06	0.37	0.14	–0.18	–0.01	0.14	0.04

Source: Dimson and Spaenjers (2014).

returns to these other collectible assets are correspondingly low (each with real returns of about 2% over the 20th century), even on a pre-transaction cost basis. Annual correlations between art, collectibles, and traditional asset classes are low (< 0.4), but there are substantial issues with measurement error (illiquidity, etc.) that may artificially reduce correlations. Details are presented in Exhibit 21.9.

21.4 R&D AND PATENTS

Research and development (R&D) and patents provide important insights into intellectual property (IP) in the context of the establishment and preservation of property rights. Unlike tangible assets, for which property rights are typically indicated by possession and usually clearly established, IP often raises challenges regarding its potential nonexcludability. This section overviews purchasing, financing, contracting, litigating, exiting, and other issues related to investment in patents.

21.4.1 Estimating Returns to R&D

Economists have devoted substantial effort to theorizing and measuring the returns to R&D expenditures. There are several complications that immediately arise. First, R&D is usually bundled with other assets inside an operating firm. Further, much ambiguity is introduced by the fact that R&D expenditure is typically undertaken over many years and therefore represents an accumulated stock of knowledge against which economic return—a flow variable—must be measured.

The theoretical and empirical evidence on the private returns (captured by specific firms) and social returns (captured by all market participants from spillover effects) to R&D is well described in a survey by Hall, Mairesse, and Mohnen (2009). The survey evidence indicates that the private returns to R&D are positive and greater than those for other forms of capital investment, and that the social returns to R&D are higher than the private returns.

The returns to patents, which are a subset of R&D assets, tend to be highly skewed, with the bulk of patents having minimal ex post value (Giummo 2014).

21.4.2 Accessing R&D through Patents

Investors have historically accessed the returns to R&D through private or public equity investments in operating entities. However, to the extent that patents or other protected IP represent the crystallization of prior R&D, ownership of patents may represent a mechanism for accessing the benefits of R&D without bearing the operational risk associated with broader investments in companies that own such IP. Investments in patents can take multiple forms, such as direct acquisition or indirect acquisition through firms or funds that specialize in the acquisition and monetization of IP.

Five key strategies for acquisition of and exit from (monetizing) patent-related IP are:

1. Acquisition and licensing
2. Enforcement and litigation
3. Sale license-back
4. Lending strategies
5. Sales and pooling

21.4.3 Patent Acquisition and Licensing Strategies

Acquisition and licensing strategies are generally built around agreements regarding royalty streams. Examples of key terms between the licensor/grantor and the licensee include:

- Minimum royalty provision: If the royalties do not hit the contracted amount within a specified commercialization period, the licensor may either terminate the license or make the license nonexclusive.
- Field-of-use provision: A licensor may grant an exclusive license for a geographical region or a particular market.
- Reservation of rights provision: The grantor may make use of the patent, most often for noncommercial research uses.

- Improvement provisions: These are provisions dealing with improvements to the patent whereby a more efficient method is created (but the new method would arguably infringe on the claims of the patent); improvements are a difficult part of the license negotiations, because either the licensor or the licensee may be the originator of the improvement.
- Audit/reporting/payment due date obligations: Licensors may want to monitor the licensee's royalty payments.
- Exclusivity responsibilities: Generally, the licensor has (sometimes limited) duties to enforce exclusivity, whereas the licensee has to report infringement cases to the licensor. This varies a great deal from license to license.

In general, license rates are typically specified as a function of revenues associated with products built on the licensed technology.

21.4.4 Patent Enforcement and Litigation Strategies

Ownership of patents may require patent enforcement and litigation to protect the value of the IP, meaning that the owner of the IP monitors[16] the use of the patent and takes legal action against those who make uncompensated or unauthorized use of the patent. In fact, an IP investment strategy can be to acquire patents or other protected intellectual property that the potential purchaser believes is being infringed on in the marketplace. This strategy has received increasing scrutiny and public debate, as "non-practicing" holders of patents seek to monetize their intellectual property portfolios.

Typically, an investor who believes that his patent is the subject of infringement will approach users of the technology and seek to negotiate a license agreement with them. This is usually far more cost-effective than litigation. However, should agreement and licensing not be achieved, the owner of the patent may seek litigation against the infringers.

While subject to risks and requiring substantial expertise, in addition to the time and costs of the litigation, actual patent litigation tends to proceed in a relatively orderly fashion, with most patent cases being resolved through settlement. For example, Janicke (2007) finds that most patent litigation (~80%) is resolved through settlement rather than trial.[17] Further, evidence is shown (see Exhibit 21.10) that settlement rates have been relatively stable through time.

The difficulty with settlements, however, is that their terms are not generally reported, so it is difficult to evaluate from public data the extent to which settlement-based outcomes generate sufficient risk-adjusted returns. However, although these outcomes are difficult to evaluate, it is known that cases resolved through trial generate median awards of $10 million—a figure generally confirmed by Mazzeo, Hillel, and Zyontz (2013).

In evaluating the returns to litigation, a key factor is the amount of time it takes to resolve a case, in part because length of time is positively correlated with costs; it takes longer to redeploy capital in new cases as old cases drag on; and, of course, there is the time value of money. The timing of resolutions is illustrated in Exhibit 21.11, with the following stylized facts:

- Defaults (represented by the first bar in the nonsettled cases 2006 chart) have the shortest time to resolution.

EXHIBIT 21.10 Outcomes of Patent Litigation

2006 Results		
ADJUDICATED		
Summary judgment	170	(7.0%)
Jury trial	52	(2.2%)
Bench trial	22	(0.9%)
Want of prosecution	31	(1.3%)
No jurisdiction	24	(0.9%)
Default	28	(1.2%)
Total adjudicated:	**327**	**(13.5%)**
SETTLED		
Consent judgment	124	(5.1%)
Voluntary dismissal	611	(25.3%)
Dismissal stating settlement	1,045	(43.3%)
Other dismissals	309	(12.8%)
Total settled	**2,089**	**(86.5%)**

Comparative Results			
Mode of Disposition	**2004: Patent Cases (2,362)**	**1986: Patent Cases (1,013)**	**1979: Patent Cases (786)**
ADJUDICATED			
Summary judgment	157 (7%)	78 (7.6%)	58 (7.6%)
Jury trial	60 (2.5%)	20 (2.0%)	17 (7.6%)
Bench trial	18 (0.7%)	53 (5%)	66 (7.6%)
Want of prosecution	38 (1.6%)	16	14
No jurisdiction	34 (1.4%)	10	8
Default	26 (1.1%)	23	2
Total adjudicated:	**333 (14%)**	**190 (19%)**	**165 (21%)**
SETTLED			
Consent judgment	144 (6%)	150	140
Voluntary dismissal	617 (26%)	673 (All agreed dismissals included)	481 (All agreed dismissals included)
Dismissal stating settlement	931 (39%)		
Other dismissals	337 (14%)	Included	Included
Total settled	**2029 (86%)**	**823 (81%)**	**621 (79%)**

Source: Janicke (2007).

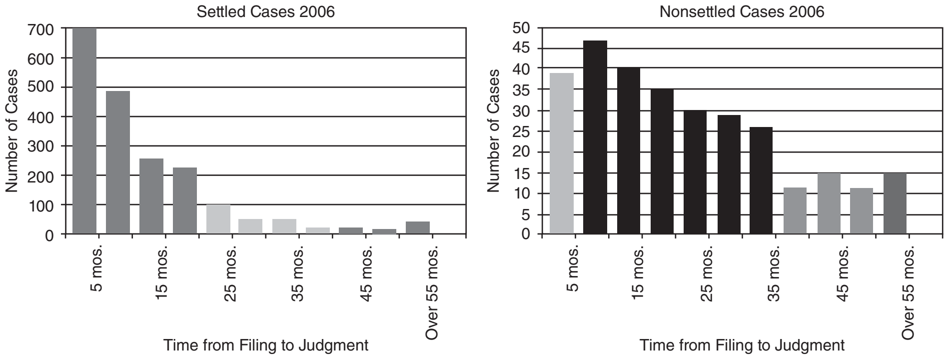


EXHIBIT 21.11 Time to Resolution for Settled Cases versus Nonsettled Cases, 2006
Source: Janicke (2007).

- Summary judgments (represented by the next six bars in the chart) range from 5 months to 35 months in duration.
- Trials (represented by bars 8 to 10 in the same chart) generally take between 35 and 50 months to resolve.
- Late dispositions (shown in the final bar on the chart) take the most time to resolve: upwards of 50 months.

21.4.5 Patent Sale License-Back Strategies

In a strategy that parallels the sale leaseback transactions of the corporate and real estate worlds, the **patent sale license-back (SLB) strategy** is in use when the patent holder sells one or more patents to a buyer, who then licenses those patents back to the original holder. In doing so, a patent seller is benefiting from the ability to monetize a portion of the intangible assets. The patent buyer then places the patent in a pool of similar technologies for out-licensing to other parties. Often, the patent buyer will participate in the licensing revenue from new licensees. By allowing the patent to be pooled with other patents, the patent owner can benefit from revenue participation generated from the potential synergies of the pooled patents.

There is also a potentially substantial tax benefit if a company lends a patent to an IP holding company in a jurisdiction with a lower tax regime than that of the previous patent holder. However, it is important to note that SLBs can incur structural problems. A borrower who has transferred title of a patent may have difficulty bringing infringement actions.

21.4.6 Patent Lending Strategies

Lending strategies backed by patents are typically separated into two classes of transactions, depending on the quality of the underlying IP:

1. Securitization: Lending backed by IP collateral; this allows separation of the IP owner's credit risk from the risk of holding the IP through the bankruptcy process.

2. Mezzanine IP lending: Lending secured by IP collateral; this usually includes warrants or other upside.

Fischer and Ringler (2014) discuss the use of patents as collateral in debt financings, and find that actual collateralizations are driven primarily by the direct economic value of the patent rather than by strategic considerations, such as the ability to potentially exclude other parties from using technology in the case of liquidations.

21.4.7 Patent Sales and Pooling

Patent owners seeking to divest patents must find buyers. Traditionally, patent buyers have entered the market for one of three reasons:

1. To purchase patents for operational use
2. To purchase patents to use as "trading cards"
3. To purchase a patent for strategic use; in this scenario, the purchaser may use the patent for defensive protection in negotiating with patent dealers.

A fourth (and emerging) class of patent buyers is made up of IP asset managers looking to buy patents for monetary exploitation.

Patent pooling, in which multiple owners of related patents agree to jointly license a number of patents to external users, is more complex than in-house licensing because of the need to divide royalty income based on revenue-sharing formulas. This can be a practical solution in industries with set standards and large quantities of patented technologies.

Two fairly recent patent pools that were highly effective at setting industry standards were the Moving Picture Experts Group (MPEG) patent pools and the DVD patent pools. Even though multiple pools had to be formed (for different MPEG formats and different DVD formats), it meant that licensees dealt with only one of a couple of pools rather than a myriad of individual companies. This simplification led to the success of both technologies. Gallini (2014) offers a much more thorough discussion of the strategic elements and consequences of patent pooling in the technology and biological sectors.

21.4.8 Risks to Investment in Patents

While there are many strategies involving patent assets, there are also many risks:

- Illiquidity. IP assets are highly illiquid assets, which often cannot be easily monetized.
- Technology/operational risk. For investors buying cash flow streams generated by IP or purchasing debt collateralized by IP, technological risk and operational risk (which may limit the investors' ability to capitalize on the IP) are major concerns; cash flows depend on successful operation of the asset, particularly when the asset is prone to heavy competitive pressure (e.g., brands or technology in a fast-moving space).
- Obsolescence. If new technology displaces current IP, the asset may be rendered worthless.

- Macroeconomic/sector risk. If macroeconomic or sector-specific factors drive down an industry, this can have significant effects on the value of a patent or a company's ability to produce cash flows from the patent.
- Regulatory risk. IP represents government-issued rights; at any point, the government could change the structure of IP authority or impose regulation on licensing/sales activities.
- Legal risk. IP transactions require a thorough understanding of IP law; failure to account for all legal implications of a transaction could result in a loss of IP value.
- Expiration risk. A patent's life is 20 years (with some exceptions for extensions, primarily in the pharmaceutical space).

21.5 INTELLECTUAL PROPERTY AND SIX CHARACTERISTICS OF REAL ASSETS

A number of investment managers have advanced the idea that certain intellectual property assets have properties that characterize them as "real assets." See, for example, Ross (2009), Auer et al. (2012), Von Bismarck et al. (2011), Fortress Investment Group (2010), and Martin (2010). Recent changes to U.S. gross domestic product (GDP) accounting, which allow certain intellectual property assets to be treated as fixed investments, are consistent with this idea (Bureau of Economic Analysis 2013). We can compare the characteristics of typical intellectual property assets to six characteristics that have been identified as being common among real assets (Martin 2014).

1. Carry low operating risk. Investing in proven and established intellectual property, such as established pharmaceutical technology, owning existing media assets rather than funding the creation of new assets (such as movie slates), and using established patents or technological IP that is already in use are lower-risk ways to invest in patents. As with traditional tangible assets, these assets may derive their value from their use in established economic processes and have value that is largely transferable from one owner to another, with less emphasis on strategic complementarities and hence niche value. The preservation of value under transfer of ownership or control is particularly emphasized in IP assets that are subject to license and sale (such as technology patents or pharmaceutical IP). Ownership rights in IP are long lasting and used repeatedly in the production process.
2. Have a positive correlation with U.S. or European inflation. Martin (2010) suggests that there is no clear evidence on the correlation of intellectual property assets and inflation. This state of uncertainty also pertains to certain other assets that have been classified as real, such as real estate. It is likely that "real" intellectual property assets have low correlation with inflation and represent a diversifier of inflation risk, unlike other traditional assets in institutional portfolios, such as equities and nominal bonds.
3. Preserve value in periods of macroeconomic instability. In general, excluding some industries such as technology and biotech, intellectual property products have low beta to the overall market. For example, we use Kenneth

French's data to calculate five-year rolling betas to the Fama-French market factor and find that the drug sector has beta in the lower half of all sectors, and that it is generally in the lowest quartile of sector betas. Elevated intermediate cash flows serve to reduce risk generally while also reducing the asset's sensitivity to exit risk in a temporarily unfavorable market environment.

4. Benefit from the scarcity of inputs in sectors like energy, manufacturing, and agriculture. It is unlikely that IP assets have any correlation with scarcity of inputs in these sectors; therefore, they do not meet this test.
5. Are essential parts of economic infrastructure. Intellectual property, and in particular "intellectual property products," as defined by the Bureau of Economic Analysis in its definition of GDP, is of growing importance. The 2013 revision to GDP to include fixed investment in long-lived intellectual property assets increased estimated U.S. GDP by $560 billion. As the 2013 changes to the system of accounts behind the calculation of GDP indicate, long-lived intellectual property products are considered a significant part of the U.S. GDP.
6. Offer long-term risk and return properties suitable for funding long-term liabilities. The focus on intellectual property products with low operating risk, ready transferability or license, and long lives provides a basis for the generation of relatively stable cash flows that may be suitable for funding long-term liabilities.

21.6 CONCLUSION

This chapter has reviewed three primary forms of IP that can serve as the basis for direct investments offering new sources of return, as well as other benefits for an investment portfolio. In the case of film production and distribution, profitability differs according to attributes such as genre, but may generally be viewed as offering return distributions skewed to the right, similar to venture capital returns. Art provides a long and somewhat plentiful history of transaction data from which estimation of historical risk and return is possible. Art has offered relatively low returns with moderate levels of risk and is subject to high transaction costs. R&D and patents are emerging as stand-alone investments, potentially offering high returns but requiring expertise in evaluation of the underlying assets.

NOTES

1. We acknowledge the coauthorship of George Martin, who derived portions of this document from Martin (2010) as well as from Wood Creek Capital Management (2008, 2011). His colleagues at Wood Creek, especially Jon Rotolo and Brian Egan, shared ongoing insights, while personifying the corporate commitment that Wood Creek has to thought leadership in investment in intellectual property and other asset classes.
2. These high correlations may be the result of nonrobustness of the Pearson correlation to outliers.
3. In our analysis and discussion, we abstract from nonmonetary benefits or only indirectly monetary benefits associated with the ownership of IP—in particular, works of art. See Geman and Velez (2015) for a discussion of this aspect to art ownership. Also see Bourdieu (1984) for a broader discussion of the interaction between nonmonetary and monetary benefits associated with owning art.

4. There are whole schools of economic thought (e.g., new growth theory) that emphasize the central role of intangibles and, in particular, their nonconvexities in economic growth. This is in contrast to traditional economic theory, which sees most production technologies as suffering from declining marginal returns to scale. See Cameron (1998) for a survey.
5. For example, the Congressional Budget Office (2005) examines results on the significant private and social returns to corporate R&D spending; Lev, Radhakrishnan, and Ciftci (2006) find that the returns to R&D leaders (defined as greater than average R&D to sales for firms in a given industry) are superior (i.e., greater in magnitude and lower in volatility) than returns to R&D followers; Hall, Jaffe, and Trajtenberg (2005) study the valuation of corporate patents; Bessen (2008, 2009) estimates the value of patents by owner and patent characteristics; Arora, Ceccagnoli, and Cohen (2008) and Chen and Chang (2010) examine the interaction between patent value and R&D; Goodwin and Ahmed (2006) document the increasing importance of intangible assets relative to earnings in the market valuation of firms; Fazzari, Brown, and Petersen (2009) find that corporate R&D spending is sensitive to available sources of corporate finance (retained earnings, debt, and equity) and the attendant frictions, which are more acute for smaller firms and younger firms, suggesting, along with the evidence presented in Hall and Lerner (2009), that IP in production by smaller firms may offer a richer source of returns to the strategic investor; and Barlevy (2007) and Rafferty and Funk (2008) find that R&D spending is procyclical. For a perspective on European intangible assets, see Sandner (2009).
6. Broad survey information on corporate R&D is published by the National Science Foundation's Survey of Industrial Research and Development.
7. See Congressional Budget Office (2005). It is worth noting that a divergence in the private returns and social returns to R&D suggests an investment opportunity regarding the social returns. Capturing such an opportunity can be done, for example, through increasing the enforcement and associated licensing of IP rights for assets with broad applicability.
8. The MPAA ceased publishing per-film costs as of 2009.
9. The transparency in DBOR numbers is high enough that substantial efforts have been made to create a futures market in DBOR results for individual films. Most recently, in the summer of 2010, efforts by the Cantor Exchange to list and trade these futures were stymied by Congress, after being approved by the Commodity Futures Trading Commission (CFTC). For more information, see www.cantorexchange.com/getdoc/85d6d736-f286-434e-b96e-6998d41ee0f8/Box-Office.
10. There are many reasons that this distribution has been adopted in the literature, including the fact that the stable Paretian distribution does not have second-order or higher-order moments and therefore mimics the extreme uncertainty about box office outcomes. As a practical matter, the absence of second moments means that the variance of an increasingly large sample of outcomes will not converge to a specific value. Secondarily, an important reason for the use of this distributional model is the availability of software to estimate stable Paretian models.
11. This is evidenced by the fact that most reported results of estimating stable Paretian distributions for box office–related revenue have $\beta = 1$, its upper boundary. See Martin (2005) for more discussion on the shape limits of the stable Paretian distribution, as well as the difficulty in estimating β precisely.
12. See De Vany and Walls (2004) for other careful evidence that suggests that break-even probabilities on films are substantially greater than the stylized 20% figure.
13. Pitt was senior economist for licensing and member services analysis at ASCAP, one of the largest performing rights organizations at the time of writing, and therefore had access to nonpublic information regarding royalty and other revenue streams.
14. While art auctions are themselves very public, the extent to which individual art transactions are brought to market and the mechanisms for preventing transactions at adverse, but market-clearing, prices are not always clear. There has also been a significant history of

collusion between auction houses in the setting of commissions, which can net the auction house up to 25% of the auction (hammer) price.

15. Ginsburgh, Mei, and Moses (2006) offer a broad overview of these methods as applied to art.
16. Descriptive information in this section is largely derived or adapted from Wood Creek Capital Management (2008).
17. Other references on outcomes in patent infringement cases include Henry and Turner (2013) and Allison, Lemley, and Schwartz (2014, 2015).

REFERENCES

Allison, J., M. Lemley, and D. Schwartz. 2014. "Understanding the Realities of Modern Patent Litigation." *Texas Law Review* 92:1769–1801.

———. 2015. "Our Divided Patent System." *University of Chicago Law Review* 82 (3): 1073–154.

Anderson, R. 1974. "Paintings as an Investment." *Economic Inquiry* 12 (1): 13–26.

Arora, A., M. Ceccagnoli, and W. Cohen. 2008. "R&D and the Patent Premium." *International Journal of Industrial Organization* 26 (5): 1153–79.

Ashenfelter, O., and K. Graddy. 2003. "Auctions and the Price of Art." *Journal of Economic Literature* 41 (3): 763–86.

———. 2006. "Art Auctions: A Survey of Empirical Studies." In *Handbook of the Economics of Art and Culture*, edited by V. Ginsburgh and D. Throsby. Amsterdam: Elsevier.

Auer, J., E. Heymann, J. Möbert, C. Schaffnit-Chatterjee, and A. Stobbe. 2012. "Real Assets: A Sought After Investment Class in Times of Crisis." Deutsche Bank Research.

Barlevy, G. 2007. "On the Cyclicality of Research and Development." *American Economic Review* 97 (4): 1131–64.

Barre, M., S. Docclo, and V. Ginsburgh. 1996. "Returns of Impressionist, Modern and Contemporary European Paintings 1962–1991." *Annales d'Économie et de Statistique*, 143–81.

Baumol, W. 1986. "Unnatural Value: Or Art Investment as Floating Crap Game." *American Economic Review Papers Proceedings* 76 (2): 10–14.

Bessen, J. 2008. "The Value of U.S. Patents by Owner and Patent Characteristics." *Research Policy* 37 (5): 932–45.

———. 2009. "Estimates of Patent Rents from Firm Market Value." *Research Policy* 38 (10): 1604–16.

Bourdieu, P. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Buelens, N., and V. Ginsburgh. 1993. "Revisiting Baumol's 'Art as Floating Crap Game.'" *European Economic Review* 37 (7): 1351–71.

Bureau of Economic Analysis. 2013. "BEA Expands Coverage of Intellectual Property Products."

Cameron, G. 1998. "Innovation and Growth: A Survey of the Empirical Evidence." Monograph. London School of Economics and Political Science.

Capgemini. 2011. "World Wealth Report."

Capgemini and RBC Wealth Management. 2013. "World Wealth Report."

Chanel, O., L. Gerard-Varet, and V. Ginsburgh. 1996. "The Relevance of Hedonic Price Indices." *Journal of Cultural Economics* 20 (1): 1–24.

Chen, Y., and K. Chang. 2010. "The Relationship between a Firm's Patent Quality and Its Market Value—The Case of US Pharmaceutical Industry." *Technological Forecasting & Social Change* 77:20–33.

Congressional Budget Office (CBO). 2005. "Background Paper: R&D and Productivity Growth."

Czujack, C. 1997. "Picasso Paintings at Auction, 1963–1994." *Journal of Cultural Economics* 21 (3): 229–47.

De Vany, A., and W. Walls. 1999. "Uncertainty in the Movie Industry: Does Star Power Reduce the Terror at the Box Office?" *Journal of Cultural Economics* 23 (4): 285–318.

———. 2004. "Motion Picture Profit, the Stable Paretian Hypothesis, and the Curse of the Superstar." *Journal of Economic Dynamics and Control* 28:1035–57.

Dimson, E., and C. Spaenjers. 2014. "The Investment Performance of Art and Other Collectibles." In *Risk and Uncertainty in the Art World*, edited by A. M. Dempster, 219–38. London: Bloomsbury.

Fazzari, S., J. Brown, and B. Petersen. 2009. "Financing Innovation and Growth: Cash Flows, External Equity, and the 1990s R&D Boom." *Journal of Finance* 64 (1): 151–85.

Fischer, T., and P. Ringler. 2014. "What Patents Are Used as Collateral? An Empirical Analysis of Patent Reassignment Data." *Journal of Business Venturing* 29 (5): 633–50.

Forsyth, R. 2012. "Art for Art's Sake? Or to Protect Wealth?" *Barron's*, May 1.

Fortress Investment Group, LLC. 2010. *Form 10-K 2010.* Retrieved from SEC EDGAR website at www.sec.gov/edgar.shtml.

Frey, B. S., and W. W. Pommenihne. 1989. *Muses and Markets, Explorations in the Economics of the Arts*. Oxford, UK: Blackwell Publishing.

Gallini, N. 2014. "Cooperating with Competitors: Patent Pooling and Choice of a New Standard." *International Journal of Industrial Organization* 36 (C): 4–21.

Geman, H., and T. Velez. 2015. "On Rarity Premium and Ownership Yield in Art." *Journal of Alternative Investments* 18 (1): 8–21.

Ginsburgh, V., J. Mei, and M. Moses. 2006. "On the Computation of Price Indices." In *Handbook of the Economics of Art and Culture*, vol. 1, edited by V. Ginsburgh and D. Throsby. Amsterdam: Elsevier.

Giummo, J. 2014. "An Examination of the Intertemporal Returns of Patented Inventions." *Research Policy* 43 (8): 1312–19.

Goetzmann, W. 1993. "Accounting for Taste: Art and Financial Markets over Three Centuries." *American Economic Review* 83 (5): 1370–76.

———. 1996. "How Costly Is the Fall from Fashion? Survivorship Bias in the Painting Market." In *Economics of the Arts: Selected Essays*, edited by V. A. Ginsburgh and P.-M. Menger, 71–83. Amsterdam: Elsevier.

Goodwin, J., and K. Ahmed. 2006. "Longitudinal Value Relevance of Earnings and Intangible Assets: Evidence from Australian Firms." *Journal of International Accounting, Auditing and Taxation* 15 (1): 72–91.

Graeser, P. 1993. "Rate of Return to Investment in American Antique Furniture." *Southern Economic Journal* 59 (4): 817–21.

Hall, B., A. Jaffe, and M. Trajtenberg. 2005. "Market Value and Patent Citations." *RAND Journal of Economics* 36 (1): 16–38.

Hall, B., and J. Lerner. 2009. "The Financing of R&D and Innovation." NBER Working Paper 15325.

Hall, B., J. Mairesse, and P. Mohnen. 2009. "Measuring the Returns to R&D." NBER Working Paper 15622.

Henry, M., and J. Turner. 2013. "Across Five Eras: Patent Validity and Infringement in United States Courts, 1929–2006." March 5. Available at SSRN: http://ssrn.com/abstract=2274383 or http://dx.doi.org/10.2139/ssrn.2274383.

Hosking, J. 1994. "The Four-Parameter Kappa Distribution." *IBM Journal of Research and Development* 38 (3): 251–58.

Janicke, P. 2007. "Patent Litigation Remedies: Some Statistical Observations." PowerPoint presentation.

John, K., S. Ravid, and J. Sunder. 2014. "Managerial Ability and Success: Evidence from the Career Paths of Film Directors." *Journal of Corporate Finance*.

Karniouchina, E., S. Carson, and W. Moore. 2010. "A Note on Revenue versus Profitability as Indicators of Motion Picture Performance." Available at SSRN: http://ssrn.com/abstract=1712088.

Lev, B., S. Radhakrishnan, and M. Ciftci. 2006. "The Stock Market Valuation of R&D Leaders." NYU Working Paper 2451/27475.

Martin, G. 2005. "The Dynamics of Alpha and Pseudo-Alpha in Hedge Fund Returns: An A-Stable Approach to Risk Endogenously Determined and Observed." Preprint.

———. 2010. "The Long-Horizon Benefits of Traditional and New Real Assets in the Institutional Portfolio." *Journal of Alternative Investments* 13 (1): 6–29.

———. 2014. "Real Assets: An Institutional Perspective." Presentation to CISDM, Amherst, MA.

Mazzeo, M., J. Hillel, and S. Zyontz. 2013. "Explaining the 'Unpredictable': An Empirical Analysis of US Patent Infringement Awards." *International Review of Law and Economics* 35: 58–72.

Mei, J., and M. Moses. 2001. "Art as an Investment and the Origin of the 'Masterpiece Effect': Evidence from 1875–2000." Mimeo, Stern School of Business, New York University.

Nakamura, L. 2009. "Intangible Assets and National Income Accounting: Measuring a Scientific Revolution." Federal Reserve Bank of Philadelphia Working Paper 09-11.

Pesando, J. 1993. "Art as an Investment: The Market for Modern Prints." *American Economic Review* 83 (2): 1075–89.

Pesando, J. E., and P. Shum. 1996. "Price Anomalies at Auction: Evidence from the Market for Modern Prints." In *Economics of the Arts: Selected Essays*, edited by V. Ginsburgh and P.-M. Menger, 113–34. Amsterdam: Elsevier.

Pitt, I. 2010. *Economic Analysis of Music Copyright: Income, Media and Performances*. New York: Springer Science & Business Media.

Rafferty, M., and M. Funk. 2008. "Asymmetric Effects of the Business Cycle on Firm-Financed R&D." *Economics of Innovation and New Technology* 17 (5): 497–510.

Ross, L. 2009. "The Rediscovery of Real Assets." Russell Investments.

Ross, M. H., and S. Zondervan. 1989. "Capital Gains and the Rate of Return on a Stradivarius." *Economic Inquiry* 27 (3): 529–40.

Sandner, P. 2009. *The Valuation of Intangible Assets: An Exploration of Patent and Trademark Portfolios*. Wiesbaden, Germany: Gabler.

Soloveichik, R. 2010. "Artistic Originals as a Capital Asset." *American Economic Review* 100 (2): 110–14.

Soloveichik, R., and D. Wasshausen. 2011. "Copyright-Protected Assets in the National Accounts." Washington, DC: U.S. National Academies of Science. Available at https://commons.georgetown.edu/blogs/copyrightnrc/files/NRCCopyright-Soloveichik-draft-5-11.pdf.

Spaenjers, C. 2010. "Returns and Fundamentals in International Art Markets." November. www.hec.unil.ch/documents/seminars/ibf/430.pdf.

Stein, J. 1977. "The Monetary Appreciation of Painting." *Journal of Political Economy* 85 (2): 1021–35.

Taylor, D., and L. Coleman. 2011. "Price Determinants of Aboriginal Art, and Its Role as an Alternative Asset Class." *Journal of Banking and Finance* 35 (6): 1519–29.

Von Bismarck, M., A. Gontownik, A. Hill, J. Hobart, J. Lerner, and I. Mendelssohn. 2011. "The Future of Long-Term Investing." World Economic Forum and Oliver Wyman.

Wood Creek Capital Management. 2008. "Intellectual Property Asset Management: Capitalizing on Innovation."

———. 2011. "Film Industry Overview."

PART

4

Commodities

CHAPTER 22

Key Concepts in Commodity Markets

Commodity investment and analysis often include a number of concepts that are unique to commodities. This is due to the fact that commodities are fundamentally different from many traditional securities. Financial securities are claims on a profit-generating enterprise, whereas commodities are simply raw materials used in the production of goods and services. Even among alternative investment strategies, commodity strategies are unique. Most alternative investments involve trading strategies based on the purchase and sale of traditional securities. For example, hedge funds typically use publicly traded stocks and bonds. The valuation methods and risk factors of private equity strategies are closely related to those of public equity strategies. Investment strategies such as those based on earnings per share (EPS) forecasting or value-based investing, which are effective when applied to traditional assets, do not apply to commodities.

The opposite is also true. Commodity strategies based on an understanding of seasonal patterns in commodity demand and market pressures leading to the existence of backwardation and contango do not translate well in traditional markets. While some of this is due to differences in the ways the commodity and traditional securities markets are organized (e.g., futures versus cash), it is also due to the fundamental difference between the economic and market factors driving traditional equity and fixed-income securities and those driving commodities.[1]

The chapter begins with a discussion of commodity spot markets and then follows up with a discussion of commodity futures and forwards markets.

22.1 ECONOMICS OF COMMODITY SPOT MARKETS

Most investing in commodities is accomplished through futures-based investments, but since futures prices are directly tied to spot prices, an understanding of commodity spot markets is critical to developing a complete understanding of commodity futures markets.

22.1.1 Commodity Prices and the Business Cycle

Although commodity prices are strongly influenced by the business cycle, a number of factors complicate any empirical analysis of the relationship. Business cycle impacts tend to vary significantly across commodity sectors and often within sectors. For example, while most energy commodities tend to respond similarly to the business cycle, agricultural commodities exhibit much greater heterogeneity in their responses to business cycle effects. These differences are related to storability and demand sensitivity to economic activity as well as exogenous factors affecting supply of some commodities. Seasonality and environmental factors such as temperature and rainfall often have larger impacts on agricultural commodities than factors related to the business cycle. It is therefore difficult to generalize the impacts of the business cycle on commodity prices. The relationships are further confounded by the impact of currency changes on commodity prices, as the U.S. dollar is still the main currency in which global trades in commodities are made, and changes in the value of the U.S. dollar versus other currencies will impact demand for some commodities. For instance, a change in U.S. monetary policy that strengthens the dollar will tend to increase the prices of major commodities in some countries, eventually reducing the demand for them. The clearest relationship between business cycles and commodity prices tends to be found in energy and industrial metals.

The relationship between commodity prices and the business cycle tends to be driven by inflation, real interest rates, and consumer and industrial demand as well as central bank policies. Of course, major technological innovations will have an impact on the marginal cost of production affecting the supply of these products.

22.1.1.1 Commodities and Interest Rates High real interest rates increase the opportunity costs of investors who hold commodities in storage, which leads to a temporary reduction in the demand for storable commodities.[2] First, high interest rates reduce the demand for storable commodities (or increase the supply) through three channels: (1) by increasing the incentive for extraction today rather than tomorrow, (2) by decreasing firms' desire to carry inventories, and (3) by encouraging investors to shift out of commodity contracts into fixed-income instruments. All three mechanisms work to reduce the market price of commodities, as happened when real interest rates were high in the early 1980s. A decrease in real interest rates has the opposite effect, lowering the cost of carrying inventories and raising commodity prices, as happened during 2002–4. Of course, other factors can alter this relationship, as we have seen a sharp decline in most commodity prices during 2014–16, when real interest rates around the globe have been at historic lows. This decline in the commodity price is attributed to maturation of the supercycle in commodities, a subject that will be discussed later in this chapter.

22.1.1.2 Commodities and Central Bank Policies The mechanism through which changes in a central bank's policies can affect commodity prices can be explained as follows. Suppose the central bank is attempting to reduce inflationary pressures in the economy and begins a contractionary monetary policy. This change will raise the real interest rate in the short run as the nominal rate increases without a corresponding increase in the inflation rate. This will lead to a decline in real commodity prices. That is, commodity prices will rise at a rate that is lower than the rate of inflation. This process will continue until commodities are undervalued enough that investors would expect a positive rate of return from appreciation in commodity prices. This positive expected return must be sufficiently high to motivate commodity firms to build their inventories even though the cost of carry is high due to high interest rates. In short, when the central bank decides to fight inflation, commodity prices tend to decline and perhaps overshoot to the downside before equilibrium is restored. Once the central bank is satisfied that inflation is under control, expansionary monetary policies are implemented, reversing the previous price declines. Economic expansion along with lower cost of carry resulting from lower real rates of interest will create a boom in commodity prices. This time, prices may overshoot to the upside.

22.1.1.3 Commodities, Inflation, and Commodity Price Patterns This scenario indicates that commodity prices follow a rather predictable pattern as a central bank's monetary policies are adjusted in response to business cycles, rising and falling as the inflation rate increases and declines. The well-documented inflation hedging characteristics of commodities may further increase their sensitivity to business cycles. A rapidly expanding economy is likely to exhibit higher inflation due to higher demand and tight labor markets. This will cause investors to seek greater exposure to commodities for their inflation hedging abilities. Increased demand will drive up commodity prices even further. Increasing commodity prices lead to an increase in inflation, prompting the central bank to tighten monetary policy, increasing real interest rates and reducing inflation. The reduction in inflation results in a reduction in demand for commodities, reducing their prices. Coincident with this hedging impact on commodity prices, there is a production demand impact. As the economy moves into an expansionary phase (particularly a strong expansion), the demand for commodities—which are direct inputs for industrial production (e.g., various forms of energy and industrial metals)—increases rapidly, resulting in large increases in their spot prices.

While demand and inflation impacts tend to dominate the pricing relationship for various forms of energy—and, to a lesser extent, industrial metals—interest rate effects tend to dominate the pricing relationship for many agricultural and livestock commodities, which may perform better in strong recessionary periods, providing a business cycle hedge rather than an inflation hedge.

22.1.2 The Properties of Spot Commodity Prices

As a foundation for commodity futures prices, this section discusses the properties of spot commodity prices.

22.1.2.1 Commodity Investing and Long-Run Returns The first question researchers tend to ask about spot commodity prices is whether the long-run price return has been positive or negative. The evidence is inconclusive.

Although it is true that prices of certain commodities have increased in the past 50 or so years, most people may not realize that the very long-term real returns on many commodities have been negative.[3] For example, from the middle of the 19th century until the end of 2005, the prices of many exhaustible resources have actually declined in real terms: aluminum (−4%), copper (−0.27%), iron (−0.56%), nickel (−1.17%), silver (−0.29%), and zinc (−0.16%). The question addressed in this section is: What should be the long-run real rate of return on exhaustible commodities? In 1931, Hotelling tried to answer this question for a commodity with a fixed quantity that is brought to market through extraction.

According to **Hotelling theory**, prices of exhaustible commodities, such as various forms of energy and metals, should increase at the prevailing interest rate—or, more specifically, the real increase in the net price of oil should increase at the real rate of interest.[4] To better understand this, consider the decision faced by the owner of an oil field. The owner can leave the oil in the ground indefinitely or extract and sell it right away and then purchase a financial asset with the proceeds. In other words, the owner can keep the oil as a physical asset or turn it into a financial asset. Suppose the price of one barrel of oil is currently \$100. If the oil is extracted and sold right away, the owner will have gained, at the end of the year, \$100 (per barrel) plus the interest earned on the \$100. By contrast, if the oil is left in the ground until a future year, the owner will get the price that is expected to prevail at that time. Which action will yield more profit? The answer is that competition and the resulting equilibrium in the oil market will cause the owner to become indifferent between the two alternatives. This will happen if the net price of oil increases at the prevailing rate of interest plus a premium to compensate the owner for the risks associated with keeping the oil in the ground. The logic behind this simple and ingenious result is straightforward and draws from the profit-maximizing behavior of the owner described in the example extended to include many resource owners.

Although Hotelling's argument does not apply to agricultural commodities whose supplies are not exhaustible, it does suggest that the long-run return to various forms of energy, industrial metals, and precious metals should be equal to the long-term interest rate. However, this theory is not without controversy. Since its publication, numerous research papers have suggested that the empirical results are not consistent with the theory. As previously stated, prices of most exhaustible resources have not kept up with inflation for the past 100 years or more. For instance, Chari and Christiano (2014) find that spot prices of exhaustible commodities have not kept up with the Treasury rate. Similarly, Bhardwaj, Gorton, and Rouwenhorst (2015) find that over the past half century (i.e., from mid-1959 to 2014), a buy-and-hold portfolio of spot commodities has not kept pace with inflation.

Note that the results regarding long-run changes in prices of exhaustible commodities assume that the production cost remains unchanged, which is clearly not true. On the one hand, as more resources are extracted, the cost of extraction could increase if places that are more difficult are searched for commodities. On the other hand, technological changes may reduce the cost of extraction. Once Hotelling's model is adjusted for these two factors, the net effect is expected to reduce the long-term growth rate in commodity prices. One additional factor that may complicate the predictions of the model is shocks to demand for the product. That is, unpredictable changes in demand could cause commodity prices to change at rates that are substantially different from the prevailing interest rate.

22.1.2.2 Commodity Prices and Supercycles As previously discussed, in the absence of significant demand shocks, prices of exhaustible commodities are expected to rise at a lower rate than the prevailing interest rate. This prediction appears to be supported by empirical evidence regarding the behavior of commodity prices over the past 200 years. However, this does not mean that commodity prices have not displayed much volatility during the past 200 years. On the contrary, supply shocks have created long periods of price increases followed by long periods of price declines referred to as **supercycles**.

Examination of commodity prices suggests four supercycles during 1865–2009, with each cycle lasting around 30 to 40 years.[5] Non-oil-price supercycles follow world gross domestic product (GDP), indicating that they are essentially determined by demand; causality runs in the opposite direction for oil prices. That is, supply shocks tend to affect oil prices, which then affect world GDP. The mean of each supercycle of non-oil commodities is generally lower than that for the previous cycle. Tropical agriculture experienced the strongest and steepest long-term downward trend through the 20th century, followed by nontropical agriculture and metals, while real oil prices experienced a long-term upward trend, interrupted temporarily during the 20th century.

The 2007–9 global economic crisis was preceded by a commodity price boom that was unprecedented in its magnitude and duration. The real prices of energy and metals more than doubled in the five years from 2003 to 2008, while the real prices of food commodities increased 75%. Whereas in the former case prices reached one of the highest levels in history, in the case of agriculture it was a reversal of the strong downward trends experienced since the 1980s. In this sense, it can be said that there was a boom of mineral, not of agricultural, prices. Similar to earlier periods of high prices, the recent one came to an end when global economic growth slowed down, diminishing demand pressures on commodity prices. However, commodity prices started to recover surprisingly quickly, until they began to decline in early 2015. The remarkable strength and length of this upswing in commodity prices reflected the extraordinary resilience of growth performance of major developing countries that demanded commodities—particularly China.

22.1.2.3 Commodity Supercycles versus Short-Term Fluctuations Supercycles differ from short-term fluctuations restricted to financial factors in two ways. First, they tend to span a much longer period, with upswings of 10 to 35 years, generating 20- to 70-year complete cycles. Second, they are observed over a broad range of commodities, mostly inputs for industrial production and urban development of an emerging economy. For example, the economic growth in the United States from the late 19th through the early 20th century led to a supercycle expansion in commodity prices that was rather well sustained and prolonged. Another upswing took place during the postwar reconstruction in Europe and was further enhanced by Japanese postwar economic emergence. These two earlier supercycles in commodity prices were driven by the resurgence of demand for raw materials during the industrialization of a major economy or a group of economies. Likewise, the current phase of supercycle expansion can be attributed to rapid and sustained Chinese industrialization and urbanization. Similar to previous supercycles, the current cycles will end when the growth rates in China, India, and other emerging economies moderate.

22.2 COMMODITY TRADING FIRMS, RISKS, AND SPECULATION

Commodity trading firms trade both physical (spot) commodities and futures contracts to take advantage of mispricings and inefficiencies that may exist in commodity markets.[6]

22.2.1 Commodity Transformation

In pursuing trading profits, commodity firms are performing the economic function of transforming commodities. These transformations can be roughly grouped into three categories: transformations in space, transformations in time, and transformations in form. Spatial transformations involve the transportation of commodities from regions where they are produced (supply regions) to the places they are consumed. Therefore, **transforming commodities** refers to the processes of altering a commodity in terms of space (i.e., location), time, and form. Commodity firms serve as **agents of transformation** by performing the tasks of transforming commodities.

Firms that engage in commodities trading attempt to identify the most valuable transformations, undertake the transactions necessary to make these transformations, and engage in the physical and operational actions necessary to carry them out. The creation of value in commodities trading involves optimizing these transformations. Most commodities have to go through a transformation process before they can be used by the final consumer of the product.

For example, to transform a commodity in space, a firm might purchase a commodity in a location in which it is relatively cheap and deliver it in a location where it sells at a premium. Thus, the firm can generate a profit by providing transportation and logistics services if it can finance, insure, and transport the commodity at a cost lower than the price spread between the two locations.

In the case of a transformation in time, a firm might attempt to profit from mispricing across time by purchasing a commodity in the cash market and storing it for delivery in the future. The firm could generate a profit if the spread between the cash and forward contract prices is greater than the cost of financing the purchase of the commodity and paying for storage, insurance, and transportation.

Transforming a commodity in form involves processing a commodity from one form to another or blending the commodity into a different form. For example, a commodity trading firm might earn the crush spread by purchasing soybeans, crushing them, and then selling soybean oil and soybean meal. The firm can profit when the spread between the two commodities is wide enough to allow for all the associated costs of processing.

While commodity trading firms operate in the spot market, they typically also engage in commodity futures transactions to hedge the price risk of their physical commodity exposure. Thus, they typically have minimal direct exposure to price risk but may bear other significant risks, including basis, spread, counterparty, and liquidity risks. It is also worth noting that commodity firms vary significantly in size, sector focus, and ownership structure.

22.2.2 Seven Commodity Trading Risks

Commodity trading exposes these firms to a number of risks. An important function of commodity traders is to manage these risks, which essentially involves transferring those risks that they do not have a comparative advantage in bearing to entities that do; this allows the traders to generate value by concentrating on their core transformation activities. It is important to note that some of these seven risks can fall into more than one category:

1. FLAT PRICE RISK. **Flat price risk** refers to risk arising from fluctuations in spot commodity prices. Traditional commodity trading involves little exposure to changes in flat price risk. In the traditional commodity trading model, a firm purchases (or sells) a commodity to be transformed (e.g., transported or stored) and hedges the resulting commodity position via a derivatives transaction (e.g., the sale of futures contracts to hedge inventory in transit) until the physical position is unwound by the sale (or purchase) of the original position.
2. BASIS RISK. This risk arises because there might be a difference between the price of the commodity that is being hedged and the price of the underlying commodity of the hedge instrument (e.g., futures or option contracts). Such price differences exist because the characteristics of the hedging instrument are seldom identical to the characteristics of the physical commodity being hedged.
3. SPREAD RISK. **Spread risk** arises from positions that are exposed to relative price variations between contracts with different delivery dates. Spread risk can occur when commodity trading firms engage in spread transactions that expose them to the risk of loss. A spread trade occurs when the same commodity is bought and sold simultaneously for different delivery dates. Many commodity hedges involve a mismatch in timing that gives rise to spread risk. For instance, a firm may hedge inventory of corn in October using a futures contract that expires in December.
4. MARGIN AND VOLUME RISK. Margin in this context refers to the profit margin of a commodity merchandiser. **Margin and volume risk** occurs when the profitability of traditional commodity merchandising depends on margins between purchase and sale prices, and the volume of transactions. These two variables tend to be directly connected because both are affected by increased demand for the commodity. That is, increased demand will tend to increase the volume as well as the margin.
5. OPERATIONAL RISK. Commodity firms are subject to a variety of risks that are best characterized as operational, in the sense that they result from the failure of some operational process rather than from variations in prices or quantities. For example, a firm that transports a commodity by sea is at risk for a breakdown of a ship or a storm that delays completion of a shipment, which often results in financial penalties.
6. MARKET LIQUIDITY RISK. Commodity trading (including specifically hedging) frequently requires firms to enter and exit positions quickly. Trading risks are lower to the extent that it is possible to enter and exit without having a large adverse impact on prices. Thus, trading is less risky—and cheaper—in liquid markets.
7. FUNDING LIQUIDITY RISK. Funding liquidity risk arises from potential losses due to limits on access to financing. Traditional commodity merchandising is highly dependent on access to financing. Many transformations (e.g., shipping a cargo of oil on a very large carrier) are heavily leveraged (often 100%) against the security of the value of the commodity. A commodity trading firm deprived of the ability to finance the acquisition of commodities to transport, store, or process cannot continue to operate.

22.2.3 Speculation in Commodity Markets

The fundamental purpose of commodity futures markets is to facilitate exchange, risk transfer, and price discovery for commercial participants. Whereas much of the regulatory enforcement of traditional security markets focuses on the improper use of private information and outright fraud, commodity futures regulation often focuses on protecting commercial participants from market manipulation because of excessive position size or improper order flow. Although speculators are recognized as serving an important role in these markets, they are generally subject to restrictive position limits to ensure that their actions do not drive prices away from fundamental levels based on supply and demand. Since spot prices of many commodities directly impact social welfare, large spot price increases in commodities (particularly energy and agricultural commodities) tend to generate a great deal of attention (including concerns about price manipulation) from regulators and politicians.

Let us first consider the definition of speculation.[7] **Price speculation** in commodities is defined as the risky purchase (or sale) of a commodity-related asset with the expectation that the price of the asset will rise (or fall) to create the opportunity for a capital gain that exceeds the equilibrium compensation for bearing systematic risk. A variety of commodity-related assets are available as instruments for speculation: commodity futures contracts, shares of commodity producing companies, and stocks of commodities kept in storage facilities. Thus, a speculator might take a long position in commodity futures because it is believed that the price is more likely to rise than fall, and the speculator hopes to earn an appropriate risk-adjusted return. It is important to note that for every long futures position there is an offsetting short position, held by someone betting that the price is more likely to fall.

Not all commodity-related investments can be considered speculation. For example, the purchase or sale of commodity futures that is done to hedge against price fluctuations that, if large enough, could lead to bankruptcy is considered an investment. In addition, the purchase of commodity-related financial assets, such as futures or oil company shares, to diversify a portfolio is considered an investment. Finally, the accumulation of oil inventories by producers or industrial consumers of a commodity as a way to facilitate deliveries and reduce the risk of stock-outs should be considered an investment. However, in practice, it is often difficult or impossible to differentiate between a speculative activity and an investment. For example, mutual funds, hedge funds, and other institutions often hold futures positions as well as commodity-producing company shares and might do so either to make an unhedged bet on future prices or as a way to diversify or to hedge against other commodity-related risks. Sometimes it is possible to clearly identify a hedging activity, but more often it is not. So in most cases what we call an investment and what we call speculation are likely to be the same thing, or at least ambiguous.

While it is difficult to distinguish among motivations for purchases of commodity-related financial assets, it is not difficult to be clear about what commodity price speculation is not: a shift in fundamentals. This could include a shift in consumption demand for oil (e.g., due to unusually cold weather) or a shift in the supply of oil (e.g., because of a change in technology). A shift in fundamentals can certainly lead to a change in the price of a commodity, and it is important to distinguish that from a price change caused by speculators or investors betting on a change in price that is not already accounted for by expected shifts in demand and supply.

We must also be clear about what price or prices we are referring to. When speculation is blamed for pushing oil prices up or down, it is usually the spot price that is being referred to—that is, the price for immediate delivery. By contrast, the futures price is the market price of a futures contract for oil to be delivered at some future point in time. When speculators (or investors) buy and sell futures contracts, the futures price may change, and we will be concerned with whether and how that change can affect the spot price.

22.2.4 The Impact of Commodity Speculation on Risk

Some have argued that commodity prices and volatilities may be at times largely determined by the activities of speculators rather than changes in the fundamentals. In particular, much attention has been focused on the impact on spot prices of increased commodity futures trading because of increased popularity of commodity index–based investments. The expanded use of financial contracts and financial engineering to facilitate commodity trading is often referred to as the **financialization of commodities**.

For example, the growth of various commodity prices in 2008, particularly crude oil, was blamed by some (politicians, the general public) on the existence of speculators who increased their investment in newly created commodity indices (Masters and White 2008), while academics tended to focus on the impact of fundamental supply and demand pressures (Till 2009; Plante and Yücel 2011; Black 2009).

Commodity index investment became viable with the introduction of the Goldman Sachs Commodity Index (GSCI) in 1991. Commodity index investment experienced modest growth through the remainder of the 1990s. However, between 2003 and 2008, commodity index investment grew at a remarkable pace. Since the prices of many commodities experienced extremely large run-ups over the same time, there was a great deal of concern that the increase of commodity prices was a direct result of the demand for commodity futures resulting from the growth in commodity index investing.

While it is beyond the scope of this section to fully address this issue, it is important to highlight the following points (Harris 2008). All the data modeling and analysis done to date indicate that there is little economic evidence to demonstrate that prices are being systematically driven by speculators in these markets. Generally, the data show that (1) prices have risen sharply for many commodities that have developed neither futures markets (e.g., durum wheat, steel, iron ore, and coal) nor institutional fund investments (Minneapolis wheat and Chicago rice); (2) markets where index trading is greatest as a percentage of total open interest (live cattle and hog futures) have actually suffered from falling prices during the past year; (3) the level of speculation in the agriculture commodity and the crude oil markets has remained relatively constant in percentage terms as prices have risen; and (4) speculators such as managed money traders are both buyers and sellers in these markets. For example, data show that there are almost as many bearish funds as bullish funds in wheat and crude oil.

As mentioned throughout this chapter, futures markets differ considerably from traditional security markets. With stocks or bonds, a limited number of securities are outstanding at any point in time. In order for new funds to flow into these markets, some investors must be willing to sell their securities. In contrast, since futures contracts are derivatives, new futures positions can be generated at any time. When an investor opens a long position in a futures contract, the other side can be an existing long closing a position or a new short opening a position. Furthermore, one must consider the fact that commodity index investors do not take physical delivery of the underlying commodities. They close their futures positions prior to expiration and roll their exposure into new futures contracts with a longer time to expiration. That is not to say that futures prices cannot affect spot prices. However, if one believes that commodity index traders put upward pressure on longer-term futures from their long positions, then it makes little sense to ignore the downward pressure they might exert from their selling of short-term futures as the contracts are rolled.

The literature on whether speculation is helpful or harmful may be summarized by the G20 Study Group on Commodities (2011) as follows: "The expansion of market participants in commodity markets increases market liquidity (including in longer-term contracts), thereby accommodating the hedging needs of producers and consumers. … On the other hand … increased correlation of commodity derivatives markets and other financial markets suggests a higher risk of spillovers."

As a final point, it is a matter of debate as to whether institutional investors should be defined as speculators in commodity markets, as some believe that these investors may require the use of commodity futures to hedge their natural exposures to inflation risk.

22.3 ECONOMICS OF COMMODITY FUTURES MARKETS

When we discuss commodity investment, the primary focus tends to be on commodity futures markets for two reasons. First, the large majority of commodity investment is implemented directly or indirectly through futures and forwards. Second, and just as important, futures markets are the primary venue for price discovery in most commodity markets. However, it is important to note that exposure to commodities can be attained through a variety of vehicles, ranging from direct ownership of a physical commodity to investment in a mutual fund that focuses on commodity-based equities. These investment vehicles will be discussed in detail in Chapter 23. This section of the chapter focuses on the key concepts relating to commodity futures and forwards.

22.3.1 Theory of Storage and Convenience Yield

The theory of storage attempts to explain the relationship between spot and futures prices by analyzing what agents hold inventories. The theory, which was first introduced in the 1930s, examines the benefit of holding the physical commodity. Inventories have a productive value because they allow the owner to meet unexpected shifts in demand, avoid the cost of frequent changes in the production schedule, and eliminate disruption to the manufacturing process. In order to represent the advantages attached to the ownership of the physical commodity, economists have developed the notion of convenience yield.

All real assets have a convenience yield, which is the benefit that comes from physical possession of an asset. It is literally a measure of the convenience of having the asset available to use. It can be viewed as an embedded timing option attached to the ownership of the commodity, since the inventory allows the owner to bring the commodity to the market when the price is high and to hold it when the price is low. Alternatively, it is a measure of how much a buyer would pay to avoid the inconvenience of constantly ordering new quantities of the asset and worrying that the supply of the asset will not arrive when needed. It is an economic benefit, not a monetary benefit. For example, a bread company may have a very high convenience yield for flour if the demand for bread is high and the inventory of flour is low. Without a sufficient inventory, the company may not be able to meet an increase in sales, since there may be delays and surcharges to acquire additional flour. However, if inventory is high and demand is relatively low, the convenience yield of additional flour will be quite low, or possibly almost zero. Convenience yields vary with the level of inventory. As inventory declines, the convenience yield rises, as consumers will pay more to ensure adequate supplies to operate their business. Therefore, convenience yield differs among market participants at a given time and often across time even for the same participant.

Measuring the convenience yield of a real asset is difficult because it varies among users and may be different for the same user from one day to the next. Alternatively, what can be measured using market prices is the **marginal convenience yield**, which is the convenience yield that will match buyers with sellers. Buyers who have a higher convenience yield will earn a consumer surplus. A **consumer surplus** is the difference between the highest price a buyer would be willing to pay (the buyer's reservation price) and the actual market price. If the market price is lower than the reservation price, the buyer earns a consumer surplus. Because we cannot observe the reservation price of each holder of inventories, we cannot measure the total amount of convenience yield that is earned.

22.3.2 Three Determinants of Convenience Yield

The convenience yield of a commodity is often compared to the dividend stream paid by a stock, because in both cases there is a return that is paid to the owner of the asset that is not paid to owners of derivatives based on the asset. Owners of equity futures contracts or stock options do not receive cash dividends. Similarly, owners of commodity futures contracts do not receive the convenience yield. Owners of stock options and equity futures contracts can earn implicit dividends, since the expected size and frequency of stock dividends is captured by derivatives pricing models. Similarly, owners of commodity futures contracts implicitly earn a portion of the commodity's marginal convenience yield, because the marginal convenience yield is incorporated into commodity futures pricing equations. However, comparing a convenience yield to a dividend yield overlooks some important differences between commodities and equities. Most significant is that convenience yield represents an economic benefit (which impacts pricing) and is unique to each commodity market participant, whereas dividends convey a transparent and easily ascertained monetary benefit equally to all holders of a stock.

Since convenience yield is related to the option embedded in the ownership of a commodity, three factors are believed to affect the level of convenience yield:

1. Among many factors, the level of inventories has a negative impact on the volatility of the commodity price. If inventory levels are low, demand and supply shocks will have a larger impact on prices. Therefore, convenience yield tends to be higher when inventory levels are low.
2. Unlike equity markets, in which changes in volatility and changes in prices tend to be negatively correlated, the relationship between volatility and commodity prices tends to be positive. The primary reason is that low inventory levels are associated with higher prices and higher volatility. Therefore, price increases tend to be associated with increases in convenience yield.
3. If there are no fundamental changes in the demand and supply of a commodity, an increase in futures prices tends to reduce convenience yield. An increase in futures prices will encourage a buildup of inventory, which may lead to a temporary increase in the spot price. However, once inventory has increased sufficiently, convenience yield will decline, moving the spot price back toward its previous value. Thus, a speculative increase in futures prices tends to reduce convenience yield.

22.3.3 Cost of Carry

The cost of carry is equivalent to the unrecoverable cost of purchasing and storing a commodity. As we will discuss later in this chapter, cost of carry is a fundamental determinant of the term structure of commodity futures. The cost of carry is calculated using the following formula:

$$\begin{aligned}\text{Cost of Carry} = {} & \text{Financing Cost} + \text{Storage Cost} \\ & + \text{Spoilage Cost} - \text{Convenience Yield}\end{aligned} \tag{22.1}$$

The major components of the cost of carry include:

- Financing costs. The standard assumption is that the commodity stored is fully financed, and that the financing cost is the cost of capital the firm applies to working capital.
- Storage costs. These costs include rental of storage facilities, insurance, inspections, transportation costs, and maintenance costs (e.g., cattle feed).
- Spoilage costs. This is the loss of value that may naturally occur through storage.
- Convenience yield. This is the benefit of owning the commodity.

Exhibit 22.1 displays an example of the cost of carry. In this case, the spot price per bushel of corn is \$4.25. Given the components of the cost of carry presented in the exhibit, one would be indifferent about choosing between purchasing the corn in the spot market for \$4.25 and carrying for three months, and purchasing the corn in the futures market for \$4.301 and taking delivery in three months. As noted previously, the convenience yield may vary by users of the commodity. Therefore, the futures prices of \$4.301 may not be the break-even price for a user of corn who assigns a higher value to the convenience of having the corn in storage and available for use.

EXHIBIT 22.1 Cost of Carry

Cost of Carry	Per Month	Three Months
Spot price per bushel		\$ 4.25
Financing rate	0.20%	\$ 0.026
Spoilage rate	0.165%	\$ 0.021
Convenience yield	0.20%	–\$ 0.026
Storage cost per bushel	\$0.010	\$ 0.030
Total cost of carry		\$ 0.051
Break-even futures price		\$ 4.301

APPLICATION 22.3.3

The spot price of a commodity is \$10 while its six-month futures price is \$10.12. Given that the annual financing rate is 3%, the annual spoilage rate is 2%, and the storage cost per month is \$0.02, what is the implied annual convenience yield?

Use Equation 22.1 to find the cost of carry noting that the futures price of \$10.12 must equal the spot price of \$10.00 plus the cost of carry. The six-month financing cost is \$10.00 × 0.03 × 6/12, or \$0.15; the six-month spoilage cost is \$10.00 × 0.02 × 6/12, or \$0.10; and the storage cost (\$0.02 × 6) is \$0.12. The convenience yield (CY) must satisfy the following equation:

$$\text{Futures Price} = \$10.12 = \$10.00 + \$0.15 + \$0.10 + \$0.12 - (\text{CY} \times \$10.00 \times 6/12)$$

$$\text{Convenience Yield} = 5\% \text{ per year}$$

Note that the computations are illustrated with simple interest rather than compounded interest for simplicity and because of the relatively short period of time. Of course, any one of the variables in the relationship could be solved given the values of all the other variables.

22.3.4 Arbitrage and the Cost of Carry without Convenience Yield

Suppose a bushel of corn will be needed three months in the future. It would theoretically make no difference to buyers whether they paid \$4.25 today in addition to \$0.31525 in cost of carry or paid a futures price of \$4.56525 for delivery in three months' time, ignoring convenience yield, commissions, and transaction costs. This break-even point is depicted in Equation 22.2:

$$\text{Break-Even Futures Price} = \text{Spot} + \text{Cost of Carry} \tag{22.2}$$

If the futures price were higher than \$4.56525, a buyer could profit by purchasing more of the commodity than needed and simultaneously selling the additional quantity in the futures market. The simultaneous cash purchase and futures market sale of a commodity to earn a riskless profit is known as **cash-and-carry arbitrage**.

If the futures price is below \$4.56525, the buyer could choose to use the futures market for delivery or purchase the commodity today. The appropriate election would depend on the buyer's convenience yield. If the buyer's convenience yield were \$0.10 per month, then the break-even futures price would be \$4.26525 (the spot price plus carry costs minus convenience yield). When the futures price is below break-even, a **reverse cash-and-carry arbitrage** transaction should be conducted to generate a riskless profit in which an initial short position in the spot market is combined with a corresponding long position in the mispriced futures contract.

As with the convenience yield, other costs of carry for a commodity vary from user to user. The cost of carry also depends on seasonal factors and on the amount of the commodity in storage at a particular time. For example, when crude oil stocks are low, the cost of storage is relatively low, since there is ample capacity in storage facilities. Conversely, when stocks of oil are high, the cost of storing increases, as storage capacity is in scarce supply.

22.3.5 Arbitrage and the Cost of Carry with Convenience Yield

We previously discussed the impact of speculation on commodity prices. Noting that most speculative transactions take place through futures markets, the cost-of-carry model allows us to examine the impact of changes in futures prices on spot prices. Note that the cost of carry may include convenience yield as depicted in Equation 22.3:

$$\text{Futures Price} = \text{Spot Price} + \text{Costs} - \text{Convenience Yield} \tag{22.3}$$

Here we have combined financing costs, storage costs, and spoilage costs into one item and described that combined item as "Costs." Written differently, the spot price can be expressed as

$$\text{Spot} = \text{Futures} + \text{Convenience Yield} - \text{Costs} \tag{22.4}$$

Now suppose that there is an increase in the futures price not due to any change in the fundamentals but due to speculative bets by a group of investors. The increase in the futures price will have no impact on the spot price if there is a corresponding decline in the convenience yield. The process of a decline in the convenience yield will depend on an increase in inventory level. Therefore, there is likely to be a temporary increase in the spot price. However, once inventory is built up, the convenience yield will decline, moving the spot price back toward where it started.

22.4 THEORIES OF COMMODITY FORWARD CURVES

The price of a commodity for delivery in the future can be higher or lower than the price of that commodity for immediate delivery. The relationship between time-to-delivery and commodity futures contract prices is known as the **forward curve**.

22.4.1 The Slope of the Forward Curve and the Cost of Carry

When the price for delivery in the future (the future price) is higher than the price for immediate delivery (the spot price), we say that the forward curve is upward sloping. This is also termed contango, as illustrated in Exhibit 22.2. When the future price is below the spot price, we say the forward curve is downward sloping, or in backwardation.

One can use the cost-of-carry model to gain insight into why the futures curve could be upward or downward sloping. It can be seen from Equation 22.4 that the spot price will be lower than the futures price if convenience yield minus cost is negative. That is, when convenience yield is low or zero, commodities that are costly to carry will be in contango. It was previously seen that low convenience yield is associated with higher levels of inventory. Conversely, when inventory levels are low or market volatility is high, convenience yield tends to be high, and therefore the futures curve will be downward sloping.

When convenience yield is high, investors with long positions in futures markets tend to earn the yield, as owners of the commodities benefit from the ownership and

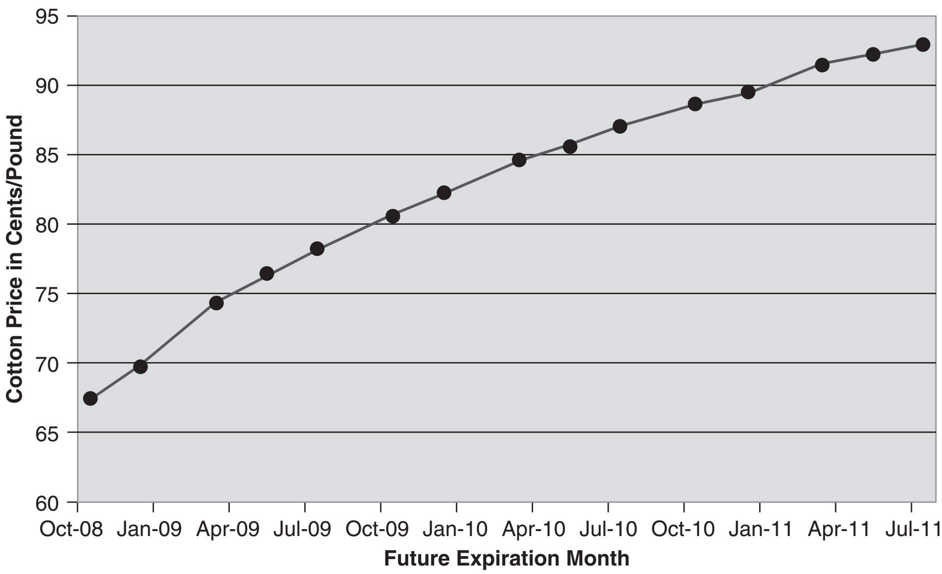


EXHIBIT 22.2 Upward-Sloping Commodity Curve, September 1, 2008
Source: Intercontinental Exchange (ICE).

can use short positions in the futures market to hedge their risks. Similarly, when markets are in contango and convenience yield is low, investors with long positions in futures markets tend to lose the storage cost, as owners of the commodity earn no benefit from holding it but have to pay the storage cost.

An important concept related to the shape of the term structure is the relationship between futures prices and expectations of future spot prices. The theories for why futures prices may be less than, equal to, or greater than expected future spot prices are similar to theories that try to explain why forward interest rates may or may not differ from expected future interest rates. In the context of commodity markets, nonspeculative or commercial accounts are typically segregated into producers and users of the commodity. In fixed-income markets, borrowers can be thought of as producers of bonds, and lenders as users of bonds. To expand on this analogy, borrowers typically have a preferred maturity date in mind when issuing bonds but are willing to shorten or extend that maturity date if market conditions dictate. In addition, a borrower may increase the quantity of bonds issued if the price is attractive. Similarly, buyers of bonds (lenders) typically consider, among other things, the prices/yields of differing maturities when making investment decisions.

Borrowers and lenders also make hedging decisions. A borrower can decide to lock in the interest rate today for a project that will be funded in the future. Alternatively, the borrower can wait until the project begins to issue bonds. These decisions are similar to a commodity producer's decision to sell future production in the forward market or to wait and sell the commodity at the future spot price.

22.4.2 Market Expectations and Forward Curves

One of the most well-known theories in economics is the **unbiased expectation hypothesis**, which proposes that current futures prices represent the market's expectations of future spot prices. This implies that forward prices are unbiased estimators of future spot prices, corresponding to a zero risk premium.

The idea behind this model is rather simple. Suppose the current futures price for delivery at the future date T is denoted by $F(t, T)$. Also, suppose the expected future spot price of the same commodity is given by $E[S(T)]$. The unbiased expectation hypothesis argues that $F(t, T) = E[S(T)]$. If the unbiased expectation holds, then returns to long or short positions in futures markets will be, on average, zero, as the future spot prices will be, on average, equal to the current futures prices. A number of empirical papers have attempted to test this hypothesis, and the results have been mixed. The main problem is that the market's expectation of the future spot price cannot be observed and has to be estimated using various proxies.

A practical trading strategy underpinned by market expectations models is **relative value arbitrage**, in which speculators who identify prices on the forward curve that deviate from their expected values either can purchase or sell those commodities outright, or can enter into spread trades by purchasing the commodity at one point on the curve and selling it at another.

For example, assume that a survey of participants in the copper market (appropriately weighted for the relative impact of each on the overall market) found that the price in one year was expected to fall by 5%. The unbiased market hypothesis would predict that the current futures price for delivery of copper in one year would be 5% below the current spot price. Any other price would violate the hypothesis. If a speculator has a different view than the market, she can take a position to profit should her predications come true. Continuing with this example, if the speculator believes that the future spot price will decline by less than 5%, she can choose to take a long position in the futures market, benefiting if the future spot price declines by less than 5%.

Note that the model assumes that markets are perfect and that, more importantly, speculators in futures markets do not demand a risk premium for taking long or short positions in futures markets. In the preceding example, the speculators who have taken a long position may earn a positive return if there is a positive risk premium in commodity markets. In this case, the futures price will have a downward bias.

Models that attempt to relate futures prices to market expectations of future spot prices have not proven to be a useful method for explaining the behavior of commodity prices through time. This does not mean that commodity markets are irrational but merely that the assumptions about perfect markets and risk neutrality are too strong for these markets.

22.4.3 Normal Backwardation and the Liquidity Preference Hypothesis

In *A Treatise on Money* (1930), John Maynard Keynes argued that commodity futures prices should typically be lower than the expected future spot prices as defined in the previous section. He defined normal backwardation as the tendency of commodity futures contracts to trade at prices below the rational expectations price. Normal backwardation suggests a possible positive risk premium for long commodity futures positions.

This argument was based on the assumption that producers of a commodity have a strong incentive to lock in a price today for future production by selling futures contracts (artificially pushing down futures prices), but that users of the commodity have a strong incentive to purchase at spot prices (artificially raising spot pricing). If there is a natural oversupply of futures contracts, then speculators will enter the market to purchase the excess supply, but only at a discount to the expected future spot price. Alternatively, discounts for future delivery may entice more users to lock in the price.

The economic rationale for this theory is that producers of a commodity have predictable production costs, which must be incurred prior to the sale, so locking in a future price for their goods is equivalent to locking in a profit margin. Users, in contrast, prefer the flexibility offered by the spot market. In the fixed-income world, this argument is similar to the liquidity preference hypothesis. The **liquidity preference hypothesis** holds that producers of bonds (borrowers) prefer long maturities, whereas consumers of bonds (lenders) prefer short maturities, distorting relative prices or rates from reflecting unbiased expectations. Producers offer attractive yields, which would mean low bond prices, to entice borrowers to extend their maturity or to induce speculators to borrow at short maturities and lend at long maturities.

Thus, speculators are compensated for providing a service to the market. They provide demand for going long futures contracts to counter the excess supply generated by net short producers, supporting futures prices at a level above where they might lie in the absence of speculators. The positive risk premium is necessary to entice speculators to enter the market by accepting avoidable downside risk.

22.4.4 Storage and Futures Curves

The impact of inventories on the convenience yield was discussed in previous sections. The storage model further expands on this and relates the relationship between spot and futures prices to inventory levels and other factors.

The relationship between spot and forward prices for a commodity depends, to a large degree, on the relationship between current storage levels and expected storage levels in the future. Storage models also consider the cost and feasibility of storing a commodity for an extended period of time and of transporting it to other locations for delivery. Storage models were introduced by Working (1949) and allow for backwardation, contango, or humped futures term structures for commodities. The spread between forward prices and spot prices is largely determined by the cost of carrying the commodity until the delivery date. As mentioned previously, convenience yield is essentially treated as a negative cost, which varies with the relative level of inventories and the likelihood of shortages, and differs for each market participant. Since convenience yield may overwhelm the cost of carry, even easily stored commodities may experience backwardation when inventories are low.

Commodities that can be either expensive or difficult to store, such as natural gas and live cattle, can have forward curves with steep positive or negative slopes. This arises from the cost-of-carry model, discussed in previous sections. In the case of natural gas, storage is feasible but becomes quite expensive as physical storage capacity limits are approached. Inadequate storage capacity for peak winter demand can result in exceptionally steep positively sloped forward curves during the fall–winter period in the natural gas futures curve (Till 2008). In comparison, live cattle are arguably nonstorable. The animal degrades after getting to market weight, resulting in a historical tendency for backwardation (Helmuth 1981).

Storage models of a commodity forward curve predict that the curve will be upward sloping when the current inventory levels are much greater than the threshold levels of demand (low convenience yield), and that it will be downward sloping when inventories are exceptionally tight (high convenience yield). Storage models are unique to real assets. They do not have a corresponding model in fixed income because, except for financing, there is no cost of carry for bonds.

Another factor incorporated into storage models is the risk of **stock-out**, which occurs when storage effectively drops to zero, resulting in consumption being entirely dependent on production and transportation networks. The risk of a stock-out typically occurs in markets with peak seasonal demand—such as natural gas or heating oil—or with annual crop cycles, such as grains. To avoid stock-out, users of a commodity have an incentive to hedge more actively at points on the forward calendar that are most susceptible to stock-out. These would be the months just before harvest for annual crops, and the later part of the heating season for natural gas and heating oil.

The theory of storage is illustrated in the **Working curve**, which positively relates the slope of the forward curve to current levels of inventory such that low inventory levels tend to be associated with a negative slope. The Working curve is typically presented as a concave, upwardly sloping curve. The nonlinearity is due to embedded real options related to inventory levels. When inventory is low, users of the commodity may face costs related to searching, delays, rush charges, and transportation. In the extreme, they may face significant costs from stock-outs as inventory is depleted, suggesting high convenience yields. At the other extreme, excessive inventory could result in negative marginal convenience yields as additional units of inventory reduce available storage capacity for competing commodities.

22.4.5 Other Models and Special Cases of Forward Curves

In certain markets, the users of a commodity rarely use the forward market to hedge future supplies. These particular markets typically exist for products that are directly consumed by the public, such as gasoline. This phenomenon is explained by the **preferred habitat hypothesis**, in which the relationship between expected spot rates and forward rates varies non-monotonically throughout the range of delivery dates due to supply and demand pressures in localized regions of the curve. The preferred habitat hypothesis is a more general version of the liquidity preference hypothesis. Whereas the liquidity preference hypothesis proposes that typical commodity producers take short positions in long-dated futures contracts in order to lock in profits, the preferred habitat hypothesis proposes that because producers have different cost structures, they may use different parts of the futures curve. Also, depending on the level of futures prices, producers might be willing to change their hedging strategies and move to a different part of the futures curve. Similarly, users may not always use front-end futures contracts to hedge their cost. As a result, the net supply or demand for specific futures contracts by natural hedgers could be positive or negative depending on market conditions and preferences of producers and users to hedge their risks. As a result, speculators may need to take short or long positions to absorb the excess demand or the excess supply of futures contracts created by hedgers. Therefore, the risk premium in the futures markets could be positive or negative depending on whether there is an excess supply or excess demand for futures contracts.

Another special case is a segmented market. Segmented markets are markets for the same product that, due to geography or other market frictions, operate relatively independently of each other; consequently, the price of the product in one market does not influence the price of the product in the other.[8] A market can be segregated across time as well. For example, the spot market for a product may be dominated by one set of buyers and sellers, and the market for delivery in the future dominated by a set of buyers and sellers with different characteristics (e.g., needs or motivations). When this happens, the forward curve for the commodity provides little useful information, as spot market players ignore what is happening in the futures market, and vice versa.[9]

In certain markets, the users of the commodity have a stronger need to hedge than the producers have. This is the opposite of normal backwardation, where it is assumed that the producer has the strongest incentive to hedge unless the future curve's shape is driven by the convenience yield. When this occurs, the forward curve would be upward sloping, as users of the commodity would need to entice sellers to trade the commodity in the forward market rather than the spot market. In such a case, the futures may trade at a premium to the expected future spot price. This relationship is referred to as normal contango. Whereas contango refers to an upward-sloping forward curve, normal contango refers to the relationship between futures prices and expected spot prices. Since we cannot directly observe the expected spot price, we can only estimate whether a market is in normal contango. The natural gas

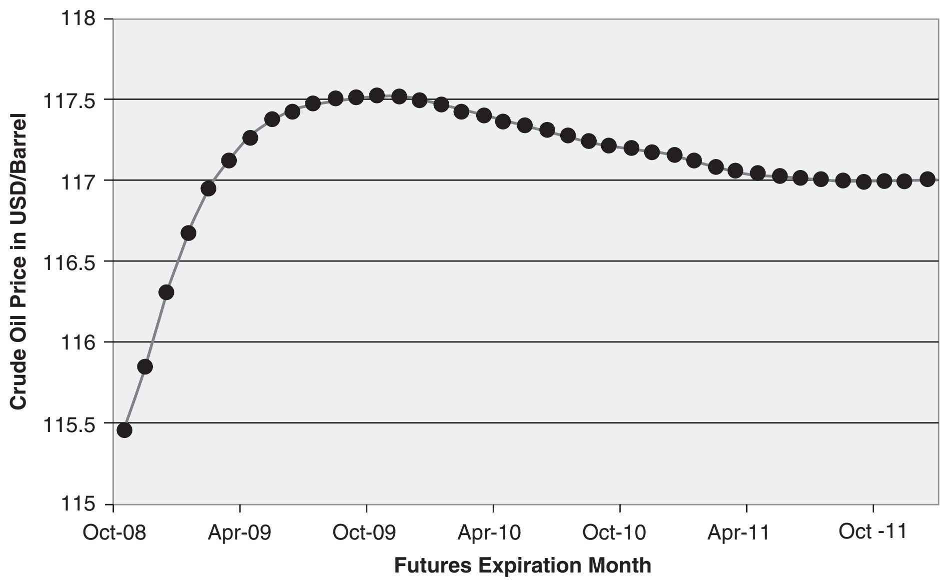


EXHIBIT 22.3 A Humped Forward Curve: NYMEX Crude Oil, September 1, 2008
Source: New York Mercantile Exchange (NYMEX).

market in the United States may exhibit this property. Utilities are believed to purchase more natural gas in the forward market than they are likely to need in order to plan for spikes in demand and reduce the risk of stock-out. When this potential demand does not materialize, the utilities sell the surplus gas in the spot market. This results in an upward-sloping futures curve.

The crude oil futures market has often exhibited a **humped curve**, which means that the market is in contango in the short term, but gives way to backwardation for longer-maturity contracts (see Exhibit 22.3).

Option-based models of the term structure focus on two types of real options embedded in commodity markets. **Real options embedded in commodity markets** are implied options involving real assets. The first type of real option is the option to extract a natural resource. A copper mine can be shut down if the price of copper falls below the marginal cost of production. Although there may be times when the spot price of copper falls below its marginal cost of production due to a temporary glut, producers will not sell forward production below cost for long; that is, they will shut down the mine. The option to extract the resource dampens the volatility of commodity prices for future delivery.

The second real option embedded in the commodity forward curve is related to inventories. Commodity markets generally have a volatility asymmetry. A **volatility asymmetry** is a difference in values between two analogous volatilities, such as is the case with commodities, in which volatility tends to be higher when prices are rising than when they are falling. This is because shortages tend to cause more problems than surpluses. The volatility asymmetry favors owning physical inventory (or short-dated futures contracts) over longer-dated futures contracts. All other things being equal, this factor will tend to flatten commodity futures curves or lead to backwardation.

22.5 DECOMPOSITION OF RETURNS TO FUTURES-BASED COMMODITY INVESTMENT

Returns on commodity futures contracts can be decomposed into three sources:

1. Spot return
2. Collateral return (or yield on risk-free securities)
3. Roll yield

The benefits of commodity investment were first explained by Anson (1998), who examined the component parts of commodity futures investment return: the spot return, roll yield, and collateral return. Examining data that spanned the years 1985–97, Anson showed that the spot return provides a diversification benefit, while the roll yield and collateral return are responsible for the bulk of a commodity investment's total return.

Spot prices increase over time, although historically they have increased less than inflation, implying that commodity spot prices alone do not serve as sufficient inflation hedges (Bhardwaj, Gorton, and Rouwenhorst 2015). Spot returns result from changes in the value of the underlying cash commodity and are generally driven by classic market factors, such as fluctuations in supply and demand for that specific asset. These factors can be the result of weather patterns or crop sizes for agricultural commodities, seasonal issues like weather or driving patterns for energy, and growth in real demand for base metals. Anson (1998) pointed out that periods of financial and economic distress can lead to market conditions that are often favorable for spot commodity prices. Because spot commodity prices tend to mean-revert over longer time horizons, spot prices cannot usually be positive sources of return over longer periods (Till 2006). As previously discussed, the Hotelling model predicts that spot prices should rise at a rate that is lower than the prevailing interest rates. In other words, the long-run returns to long positions in exhaustible commodities tend to be low, and empirical evidence tends to support this assertion.

When futures contracts are used, the **income return** (i.e., collateral yield) of a commodity investment results from the return of the cash collateral, which is usually a Treasury bill rate in the United States, although the cash collateral can be in other forms, such as Treasury Inflation-Protected Securities (TIPS), money market securities, and other liquid assets. Most commodity trading programs include a collateral feature. Whereas actual commodity futures positions require margins less than the notional value of the contracts, most commodity indices assume fully collateralized positions; thus, the amount of collateral is equal to the notional value of the futures contracts.

Because commodity investors and speculators are generally not in the market to take ownership of the actual physical commodity they are trading, a futures position needs to be closed out or rolled prior to expiration. Rolling involves selling a futures contract that is relatively close to expiration and opening a new position in a contract for the same commodity that expires at a later date. Roll return is often mistakenly thought of as a profit or loss occurring at the time of the contract roll. However, roll return (or roll yield) is defined as the portion of the return of a futures contract that is due to the change in the basis (difference between futures price and spot price) over time. Roll return actually accrues over time, from the time the investor goes long the futures contract to the time of the roll, much like a bond coupon (but can be a negative value, unlike a bond's coupon). While one would expect the roll return to a long futures position to be positive if the forward curve for the commodity slopes down (backwardation) and negative if the forward curve slopes up (contango), the roll return is also affected by changes in cost of carry (interest rates, storage costs, convenience yields, and, for financial futures, dividends). As discussed previously, the source of roll return is primarily the presence of significant convenience yield. That is, markets tend to be in backwardation when convenience yield is high, and long positions in futures contracts earn the convenience yield. While Bhardwaj, Gorton, and Rouwenhorst (2015) found that spot commodity returns have not kept up with inflation from mid-1959 to 2014, commodity futures have significantly outpaced inflation, providing returns only slightly lower than equities over the period as a result of positive collateral and roll yields.

Scarcity in commodity markets can provide a source of return to commodity investors, but the difficulty can be in determining when this market pattern is occurring. Using the forward curve, relative price differences of futures contracts across delivery months can be measured. If the forward curve is downward sloping, meaning spot prices are higher than those in the futures market, this price pattern can indicate scarcity, as a premium is being offered for the immediately deliverable commodity. This price pattern may also indicate a lack of excess of commodity inventory (Till and Eagleeye 2005).

22.6 COMMODITIES AS AN INFLATION HEDGE

According to Greer (1978), one important property of commodity investments—besides diversification—is that they can be used as a hedge against inflation.

22.6.1 Rationale of Commodities as an Inflation Hedge

The value of nominally priced assets, such as stocks and bonds, decreases when inflation increases. In theory, stocks represent claims against real assets and their future free cash flows; however, to the extent that companies have nominally fixed contracts with suppliers, customers, workers, and capital, stocks do not react directly to an increase in inflation. Stocks represent company ownership and a share in the payout of dividends (or increased proportional ownership in the case of share buybacks) or, as a terminal event, acquisition/liquidation value. Debentures represent a claim on debt repayment (or company/guarantor assets, in the case of default), and, in contrast to stocks, the debenture holder receives either a fixed or a floating stream of cash flows. The present value of the future cash flows depends on their size and timing and assumed discount rate(s). As inflation leads to increases in nominal interest rates, the discount rate increases, leading to a lower present value of cash flow—and, therefore, lower prices—for stock and bond investments.

In contrast, commodity futures prices may be related to the expected spot price in the future. Therefore, commodities can be inflation hedges, as commodity futures prices tend to increase when expected inflation increases. To the extent that commodities are part of the basket of goods from which the aggregated inflation of an economy is calculated, the increase of commodity prices itself causes inflation. However, the core inflation rate, which many central banks target for monetary policy, excludes energy and agricultural commodities due to their high volatility. Thus, the impact of commodity prices on core inflation tends to be nonlinear. That is, small moves in energy and agricultural commodity prices may have no impact on core inflation, as they are excluded from the measure of core inflation. However, large moves in these commodity prices can have significant effects on the macroeconomy, which are reflected in the core inflation rate. This is particularly true of various forms of energy (and, to a lesser extent, industrial metals), as they have an impact on consumer demand through the availability of discretionary spending and through core inflation in that they are inputs into the production of consumer goods.

22.6.2 Evidence Regarding Commodities as an Inflation Hedge

Empirical studies show that returns of commodity futures are positively correlated with changes in inflation and that commodities provide an effective inflation hedge during periods of high inflation. Bhardwaj, Gorton, and Rouwenhorst (2015) show (for the period 1959 to 2014) that commodities can be used as a hedge against inflation. They find a positive correlation between a total return index of commodity futures and the U.S. Consumer Price Index (CPI) of 0.47 for five-year averages of monthly futures. In contrast, the correlation coefficients between inflation and stocks and bonds are both negative at –0.10 and –0.20, respectively.

Although there is a general perception that commodities are effective inflation hedges, inflation characteristics vary significantly over time as well as across countries, commodities, and time horizons. Adams, Fuss, and Kaiser (2008) consider the global inflation hedging abilities of an equally weighted composite index of S&P GSCI subindices as well as each of the individual subindices. Since investors are concerned about inflation in their respective home countries, U.S. inflation may not be a relevant measure for all investors. For instance, European or Asian investors shift money into commodities when inflation in Europe or Asia rises. Adams, Fuss, and Kaiser (2008) find that the correlation between monthly U.S. inflation and commodities was 0.31 from 1983 to January 2007, largely driven by the energy subindex. In contrast, the correlations with EU and Asian inflation were 0.10 and –0.06, respectively. While these results suggest that commodities may not be an effective hedge for international inflation, some mitigation factors should be considered. First, inflation measures in those regions are averages of inflation in different countries, which might bias the estimated correlation; and second, European or Asian investors have to consider exchange rate movements when investing in dollar-denominated commodities, so the effects of exchange rate movements have to be considered as well. Furthermore, Bhardwaj, Gorton, and Rouwenhorst (2015) have suggested that correlations between commodities and inflation may be stronger for longer time horizons.

Furthermore, as suggested earlier, correlations between commodities and inflation tend to vary significantly over time. For example, Adams, Fuss, and Kaiser (2008) find that the one-year correlation coefficients are strongest for industrial metals and the composite index, and fluctuate quite dramatically (between +0.8 and –0.8) over time.

Similarly, Denson (2006) shows that the rolling correlations of commodity prices with U.S. inflation fluctuate strongly in the short run but are more stable and, on average, positive when considered over periods beyond three years. Thus, in the long run, a positive relationship between U.S. inflation and commodities exists. Furthermore, Denson decomposes inflation into expected inflation and unexpected inflation and finds that the effect of unexpected inflation is much larger than that of expected inflation, so the hedging property is much higher when inflation is unexpected, which is logical, as expectations for inflation should already be embedded in futures pricing. Erb and Harvey (2006) conclude that commodities that are storable to only a limited extent, such as heating oil and livestock, provide a better hedge against unexpected inflation than commodities that are very suitable for storage. One reason for this could be that an increase in demand for the former type of commodities increases prices directly, while in the latter case, prices are affected only after the inventory has been depleted.

22.6.3 Rationale of Financial Asset Correlations with Inflation and Commodities

An attempt to explain the inverse relationship between financial assets and both commodities and inflation is provided by Akey (2005). If it is reasonable to assume sticky output prices (dependent on relative bargaining power), an increase in inflation due to rising commodity prices raises the net costs for firms that buy commodities as inputs for production. Higher costs reduce the profits of such firms and thus put downward pressure on stock prices. Over time, higher commodity prices lead to new commodity-producing firms entering the market. This increases the supply of commodities while firms reduce their demand for commodities due to the higher costs. Both effects simultaneously combine to lower commodity prices, and as the central bank works to reduce inflation to more normal levels, profits of firms begin to improve. In turn, stock prices increase.

Another, possibly more relevant, explanation is proposed by Greer (2000). When inflation increases, the central bank is expected to respond by raising interest rates, which reduces the present value of future cash flows and thus lowers stock and bond prices. Commodity prices, by incorporating the new expected inflation rate, provide investors with an incentive to move out of stocks and bonds and into commodities.

22.7 COMMODITIES AND EXCHANGE RATES

Commodities account for a quarter of merchandise trade, which accounts for a quarter of global GDP. Since many developing countries depend on the export of only a few commodities, it is important to understand the effects of exchange rate changes on commodity prices. Many commodities are denominated in U.S. dollars, and therefore exchange rate movements vis-à-vis the dollar affect the prices for exporters and importers of commodities. Thus, in addition to inflation, interest rates, and commodity-specific risks, investors also face exchange rate risk.

However, the effects of a volatile exchange rate go beyond the investors' risk. A general depreciation of the dollar increases dollar-denominated commodity prices as commodity exporters from other countries demand a higher price in return for the exchange rate loss, while a higher price becomes equally affordable abroad, and vice versa. Exchange rate movements of single currencies can have substantial effects on the profits of commodity-producing firms as well as on supply changes. One prominent example is the case of South Africa, where the rand depreciated against the dollar by more than 35% in 2001, while the gold price in dollars actually decreased by 2.9%. This combination raised profits of South African gold companies, which expanded production in the following period. However, it should be noted that the supply of nonstorable commodities is fixed in the short run, since investment in commodity infrastructure (such as storage facilities) can take years; thus, price movements can be caused by either changes in exchange rates or changes in demand. Only in the long run do further investments in commodity production lead to an increase in supply. The short-run supply of storable commodities is somewhat more elastic as long as commodity-producing firms still have inventories.

This makes it more difficult to predict commodity price movements. On the one hand, if world demand is high in the coming years, if new investments in commodities are not yet completed, and if the U.S. dollar depreciates, commodity prices should mean-revert toward the long-term real price level, which in theory is the cost of production. On the other hand, an appreciating U.S. dollar and higher supplies would tend to cause commodity prices to decline.

As can be seen in Exhibit 22.4, correlation coefficients (linking commodity prices and USD prices) have the expected negative sign. This supports the argument that an increase in the exchange rate, which generally corresponds to an appreciation of the dollar, decreases commodity prices (and vice versa), although not all commodities are affected by the same magnitude. For example, prices of those commodities that are not heavily traded globally are not highly correlated with exchange rates.

The currencies of countries whose major exports are commodities are often referred to as **commodity currencies**. Examples of countries with commodity currencies include Australia, Brazil, Canada, New Zealand, Chile, Norway, South Africa, and Russia. These currencies have traditionally moved in close correlation with commodity prices due to underlying economic forces. For example, if the price of a commodity that is a significant portion of a country's exports declines, the balance of trade worsens, putting downward pressure on the country's currency. Thus, a decrease in the commodity price results in a weaker currency. As Kato (2012) points out, these relationships are particularly strong for countries such as Canada and

EXHIBIT 22.4 Correlations among Commodities and USD Index

1996–2015	USD Index	S&P GSCI Comm.	S&P GSCI Energy	S&P GSCI Industrial Metals	S&P GSCI Precious Metals	S&P GSCI Agriculture
USD Index	1.00	−0.41	−0.35	−0.41	−0.42	−0.35
S&P GSCI Commodities	−0.41	1.00	0.98	0.51	0.33	0.40
S&P GSCI Energy	−0.35	0.98	1.00	0.40	0.25	0.24
S&P GSCI Industrial Metals	−0.41	0.51	0.40	1.00	0.36	0.34
S&P GSCI Precious Metals	−0.42	0.33	0.25	0.36	1.00	0.29
S&P GSCI Agriculture	−0.35	0.40	0.24	0.34	0.29	1.00

Source: Bloomberg and authors' calculations.

Australia, for which raw commodities account for more than half of total exports. However, Kato finds that the financialization of commodities may have paradoxically resulted in the Australian dollar having an increased correlation with broad commodity indices such as CRB (originally Commodity Research Bureau) and S&P GSCI while concurrently having a decreased correlation with an index of the commodities that Australia actually exports. Furthermore, while Chen, Rogoff, and Rossi (2010) find that commodity currencies were effective predictors of future commodity prices using data up to 2008, Kato (2012) finds that the predictive power no longer exists post-2008. In summary, Kato argues that the financialization of commodities has resulted in two significant changes in commodity currencies. First, commodity currencies now have stronger relationships with broad commodity indices than with the prices of local commodities, and second, commodity currencies have lost their predictive power for commodity price movements.

22.8 REBALANCING AND HISTORICAL PERFORMANCE OF COMMODITY FUTURES

As mentioned earlier, collateralized commodity futures returns can be decomposed into the spot return, the collateral yield, and the roll return. Not surprisingly, even if spot returns are zero, an investor might generate positive roll returns by earning a portion of the convenience yield if a futures contract market is in backwardation. Similarly, an upward-sloping (contango) term structure might result in negative returns, although a significant collateral return would mitigate this effect.

22.8.1 Empirical Evidence on the Effect of Rebalancing on Return

Furthermore, in a commodity futures portfolio context, Greer (2000) finds that rebalancing yield can provide a significant contribution to returns. In the context of commodities, **rebalancing yield** is additional return produced by the process of rebalancing indexed commodity products. Enhanced returns can be attributed to the tendency of commodity prices to be mean-reverting while exhibiting low correlations with other commodity returns. The rebalancing yield of a fixed-weight portfolio is expected to provide a greater contribution to the returns of the portfolio when the portfolio constituents are volatile and have low correlations with each other. Erb and Harvey (2006) find that the average correlation across a set of 16 commodity futures from 1982 to 2004 was very low at about 9%, and the average volatility was relatively high at about 25%. Thus, one would expect that a commodity portfolio might achieve greater rebalancing benefits than a stock or bond portfolio, which tends to have much higher average within-class correlations.

22.8.2 The Effects of Rebalancing When Prices Do Not Mean-Revert

To understand the rebalancing yield and the impact of diversification, it is important to note that in the absence of mean reversion in prices, Chambers and Zdanowicz (2014) show that rebalancing will have no impact on the expected future value of an

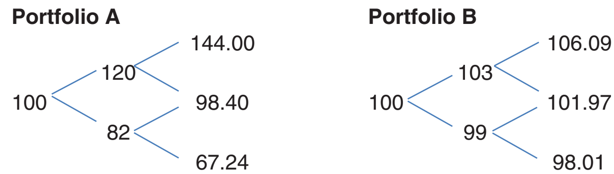


EXHIBIT 22.5 Paths of Two Portfolios

investment in a portfolio of commodities. However, rebalancing changes volatility, which in turn will have an impact on the probabilities of future growth paths of the portfolio. Specifically, different volatilities will change the probability that the future value of the portfolio will exceed a particular value. In other words, certain statistical properties of the distribution of the future value of the portfolio (excluding its mean) will change as return volatility changes.

To see this, consider two investments. Portfolio A can increase by 20% or decrease by 18% each year with equal probabilities, and Portfolio B can increase by 3% or decrease by 1% each year with equal probabilities. These values were chosen so that each portfolio has an arithmetic mean return of 1% per year. Clearly, the rate of return on Portfolio B is less volatile than the rate of return on Portfolio A. If the initial investment in both portfolios is \$100, then their paths after two periods can be seen in Exhibit 22.5.

The expected values of the two portfolios after two periods are equal to each other and equal to \$102.1, after an expected return of 1% each year, despite the difference in their volatilities. This illustrates a key point: the long-term future expected value of an investment depends directly on its arithmetic mean return, not its geometric mean return.

However, we can see that there is only a one in three chance that Portfolio A would make money after two years, while there is a two in three chance that portfolio B would make money after two years. In other words, the most likely outcome is for Portfolio A to lose money and for Portfolio B to make money. In this sense, rebalancing the commodity portfolio in order to maintain a diversified portfolio and to reduce its return volatility can help to improve its performance through its reduced risk.

22.8.3 The Effect of Rebalancing on Geometric and Arithmetic Mean Returns

Some studies point out that the geometric mean return on an investment can be written as:

$$R_C \approx \bar{R} - \frac{1}{2}\sigma^2 \tag{22.5}$$

Here, R_C is the per-period (e.g., annual) compounded rate of return, $\bar{R}$ is the per-period arithmetic mean return, and σ^2 is the per-period variance of the return. It can be seen that if the volatility of the return is reduced, the geometric mean return will increase. However, as was demonstrated in Exhibit 22.5, this will have no impact on the expected future value of the portfolio. Other statistical properties (e.g., median, mode, and skewness) of the distribution of the portfolio's future values are, of course, affected as the volatility of the per-period rate of return is changed.

22.8.4 Historical Analysis of Commodity Returns

Historically, futures-based commodity investment has provided favorable returns as well as diversification benefits. Bhardwaj, Gorton, and Rouwenhorst (2015) find that performance of spot prices and futures prices varied widely between 1959 and 2014. In some periods, roll returns were particularly high, improving the performance of futures contracts (e.g., 1959–2004). On the other hand, roll returns were low or negative for 2005–14, leading to poor performance for futures contracts in comparison to spot prices. Exhibit 22.6 displays these results. Note that the exhibit gives the (arithmetic) average annual return on an equally weighted portfolio of collateralized commodity futures that is rebalanced monthly, as well as the corresponding index of spot prices. The monthly average return is annualized by multiplying by 12. This means that the compounded monthly returns will be lower than the given figures.

The analysis of historical performance can provide useful information for forecasting future commodity investment returns, volatility, and correlations with other asset classes. However, it is important to point out that the measures of historical performance of commodity futures depend not only on the period of time considered but also on portfolio weighting and roll procedures. Since futures positions must be rolled from short-term to longer-term contracts on a regular basis, and since roll return may be a significant portion of commodity returns, the roll methodology can have a significant impact on measured returns. However, difficulty in measuring a long-run return is not unique to commodity futures markets. Commodity futures investment—like most alternative investment strategies, including hedge funds, managed futures, and private equity—has a relatively short history of clearly defined performance. For example, crude oil futures were first listed on a futures exchange in 1983. Data for energy markets prior to 1983 have been assembled by researchers from over-the-counter sources and from spot market transactions.

Furthermore, with the growth of the commodity index–based investment, many commodity markets have become financialized, or transformed from purely commercial markets into markets with a significant investor presence, leading to concerns that the historical track record may be of questionable value in projecting future

EXHIBIT 22.6 Commodity Futures and Spot Returns for Selected Periods

	Nominal		Inflation-Adjusted	
	Futures	Spot	Futures	Spot
1/2005 to 12/2014	5.09%	9.42%	2.93%	7.26%
1/1995 to 12/2004	8.55%	8.20%	6.13%	5.78%
1/1985 to 12/1994	9.69%	6.66%	6.15%	3.13%
1/1975 to 12/1984	9.01%	7.08%	1.91%	−0.02%
1/1965 to 12/1974	19.23%	13.18%	14.10%	8.07%
7/1959 to 12/1964	3.88%	2.98%	2.62%	1.71%

Source: Bhardwaj, Gorton, and Rouwenhorst (2015).

EXHIBIT 22.7 Performance of Commodities, Equities, and Bonds

	Annualized Return	Annualized Standard Deviation	Correlations: USD	Commodity Index	Stocks	Bonds
1996–2000						
USD[1]	5.3%	7.2%	1.00	−0.11	−0.02	−0.84
Commodity Index[2]	9.3%	20.4%	−0.11	1.00	0.03	0.15
Stocks[3]	10.7%	14.4%	−0.02	0.03	1.00	0.08
Bonds[4]	3.6%	5.4%	−0.84	0.15	0.08	1.00
2001–5						
USD	−3.6%	7.7%	1.00	−0.16	−0.14	−0.86
Commodity Index	9.8%	22.1%	−0.16	1.00	0.04	0.10
Stocks	0.6%	14.8%	−0.14	0.04	1.00	−0.07
Bonds	7.3%	5.9%	−0.86	0.10	−0.07	1.00
2006–10						
USD	−2.8%	9.7%	1.00	−0.59	−0.66	−0.83
Commodity Index	−5.7%	28.9%	−0.59	1.00	0.62	0.26
Stocks	0.4%	19.9%	−0.66	0.62	1.00	0.40
Bonds	6.9%	6.8%	−0.83	0.26	0.40	1.00
2011–15						
USD	4.5%	7.3%	1.00	−0.67	−0.55	−0.62
Commodity Index	−15.2%	19.9%	−0.67	1.00	0.53	0.32
Stocks	5.4%	12.9%	−0.55	0.53	1.00	0.32
Bonds	1.2%	3.9%	−0.62	0.32	0.32	1.00

[1]U.S. Dollar Index spot rate.
[2]S&P GSCI Total Return CME.
[3]MSCI World USD.
[4]Barclays Capital Bond Composite Global Index.
Source: Bloomberg and authors' calculations.

returns. In particular, much of the case for commodity investment is derived from the low levels of historical correlation between the returns on commodity futures and stock or bond investments. If the correlation is rising over time, commodities may be less diversifying than previously estimated.

There remains considerable research-based uncertainty surrounding the outcome of the financialization of commodities debate. The historical track records for commodity futures portfolios indicate attractive returns with diversification benefits. However, there is some concern that the financialization of commodities may have implications on the applicability of these historical results for predicting future returns and correlations. (See Exhibit 22.7.)

22.8.5 Historical Commodity Returns and Financialization

A number of academic and industry research papers have been written on the subject of financialization during the past 10 years. The primary motivation behind these studies is to determine if increased investments by pension funds, endowments, and hedge funds in commodity-related securities has caused the risk-return profile of commodities to change.

For example, one set of studies examines the relationship between commodity index returns and equity index returns to determine whether the relationship has changed in recent years.[10] These studies find that the extent of comovement between the two asset classes had not changed significantly from the 1980s to 2008. Furthermore, the low correlation between the two actually became negative in the last five years of the study. Even when the empirical study was restricted to periods of extreme returns, no increase in comovement between commodities and equities was found. On the other hand, more recent studies have shown that the correlation between commodity and equity markets has increased sharply since the second half of 2008. The markets seem to have recently included commodities in the category of risky assets, which are highly correlated with equities in a "risk on, risk off" volatility regime. In addition, recent studies suggest that futures-based commodity investment may no longer be an attractive diversifier for traditional stock and bond portfolios, as not only has correlation with equities increased but returns have been lower due to roll return being low since 2008.

22.9 CONCLUSION

This chapter has presented an overview of a number of key concepts in commodities markets. It began with a discussion of spot markets. While some controversy exists as to whether spot commodities should provide a reasonable long-term return, there is a great deal of evidence that suggests that spot commodity prices are strongly influenced by the business cycle. Spot price return is only one part of the return to commodity futures. Over the long term, collateralized commodity futures have provided collateral returns and roll returns, which have significantly outpaced spot returns. Thus, the term structure of commodity futures is a fundamental determinant of the returns to commodity futures–based investments.

Although commodity futures have historically provided returns similar to those of equities, their appeal in a portfolio context may be driven more by their business cycle sensitivity, low correlations with traditional investments, and inflation hedging potential. However, it is important to note that while historical returns can provide some insight into future returns and correlations of commodity investments, there is evidence that the financialization of commodity futures may impact the predictive power of historical data.

NOTES

1. The returns in commodity futures investing are due to highly nuanced factors. For background, see Till (2006) and Kazemi et al. (2007).
2. Frankel (2006).
3. Harvey, Kellard, Madsen, and Wohar (2010).
4. See Gaudet (2007) and Chari and Christiano (2014).
5. Erten and Ocampo (2013).

6. This section is based on Pirrong (2015).
7. This section draws from Knittel and Pindyck (2013).
8. Good examples of this are the natural gas markets in North America and Europe. Because the cost of shipping liquefied natural gas (LNG) from Europe to North America is high, the markets are effectively segregated. Markets for water are typically segregated across much smaller geographical regions.
9. There is reasonable empirical evidence that the long-term market for crude oil (beyond 18 months) is segmented from shorter-term delivery. See Lautier (2005) for a discussion of this research.
10. See, for example, Buyuksahin, Haigh, and Robe (2010), Chong and Miffre (2008), Kawamoto et al. (2011), and Zaremba (2015).

REFERENCES

Adams, Z., R. Fuss, and D. G. Kaiser. 2008. "Macroeconomic Determinants of Commodity Futures Returns." In *The Handbook of Commodity Investing*, edited by F. J. Fabozzi, R. Fuss, and D. G. Kaiser, 87–112. Hoboken, NJ: John Wiley & Sons.

Akey, R. P. 2005. "Commodities: A Case for Active Management." *Journal of Alternative Investments* 8 (2): 8–30.

Anson, M. 1998. "Spot Returns, Roll Yield, and Diversification with Commodity Futures." *Journal of Alternative Investments* 1 (3): 16–32.

Bhardwaj, G., G. Gorton, and G. Rouwenhorst. 2015. "Facts and Fantasies about Commodity Futures Ten Years Later." NBER Working Paper.

Black, K. 2009. "The Role of Institutional Investors in Rising Commodity Prices." *Journal of Investing* 18 (3): 21–26.

Buyuksahin, B., M. Haigh, and M. Robe. 2010. "Commodities and Equities: Ever a 'Market of One'?" *Journal of Alternative Investments* 12 (3): 76–95.

Chambers, D., and J. Zdanowicz. 2014. "The Limitations of Diversification Return." *Journal of Portfolio Management* 40 (4): 65–76.

Chari, V. V., and L. J. Christiano. 2014. "The Optimal Extraction of Exhaustible Resources." Economic Policy Paper 14-5, Federal Reserve Bank of Minneapolis.

Chen, Y.-C., K. S. Rogoff, and B. Rossi. 2010. "Can Exchange Rates Forecast Commodity Prices?" *Quarterly Journal of Economics* 125 (3): 1145–94.

Chong, J., and J. Miffre. 2008. "Conditional Return Correlations between Commodity Futures and Traditional Assets." EDHEC Working Paper.

Denson, E. 2006. "Should Passive Commodities Investments Play a Role in Your Portfolio?" Investment Viewpoints, UBS Global Asset Management.

Erb, C. B., and C. R. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal* 62 (2): 69–97.

Erten, B., and J. A. Ocampo. 2013. "Super Cycles of Commodity Prices Since the Mid-Nineteenth Century." *World Development* 44 (C): 14–30.

Frankel, J. 2006. "Commodity Prices, Monetary Policy, and Currency Regimes." NBER Working Paper C0011.

G20 Study Group on Commodities. 2011. "Report of the G20 Study Group on Commodities under the Chairmanship of Mr. Hiroshi Nakaso." November.

Gaudet, G. 2007. "Natural Resource Economics under the Rule of Hotelling." *Canadian Journal of Economics/Revue Canadienne d'Économique* 40 (4): 1033–59.

Greer, R. J. 1978. "Conservative Commodities: A Key Inflation Hedge." *Journal of Portfolio Management* 4:26–29.

———. 2000. "The Nature of Commodity Index Returns." *Journal of Alternative Investments* 3 (1): 45–52.

Harris, J. 2008. "Written Testimony of Jeffrey Harris, Chief Economist Before the Senate Committee on Homeland Security and Governmental Affairs". May 20. www.cftc.gov/idc/groups/public/@newsroom/documents/speechandtestimony/oeajeffharristestimony-052008.pdf.

Harvey, D. I., N. M. Kellard, J. B. Madsen, and M. E. Wohar. 2010. "The Prebisch-Singer Hypothesis: Four Centuries of Evidence." *Review of Economics and Statistics* 92 (2): 367–77.

Helmuth, J. W. 1981. "A Report on the Systematic Bias in Live Cattle Futures Prices." *Journal of Futures Markets* 1 (3): 347–58.

Hotelling, H. 1931. "The Economics of Exhaustible Resources." *Journal of Political Economy* 39 (2): 137–75.

Kato, H. 2012. "Changes in the Relationship between Currencies and Commodities." Bank of Japan Review Series, 2012-E-2, March.

Kawamoto, T., T. Kimura, K. Morishita, and M. Higashi. 2011. "What Has Caused the Surge in Global Commodities Prices and Strengthened Cross-Market Linkages?" Bank of Japan Working Paper Series, 11-E-3, May.

Kazemi, H., T. Schneeweis, R. Spurgin, and G. Martin. 2007. "Real Assets in Institutional Portfolios: The Role of Commodities." Alternative Investment Analytics LLC Research Report.

Keynes, J. M. 1930. *A Treatise on Money*. London: Macmillan.

Knittel, C. R., and R. S. Pindyck. 2013. "The Simple Economics of Commodity Price Speculation." MIT Sloan School of Management.

Lautier, D. 2005. "Term Structure Models of Commodity Prices: A Review." *Journal of Alternative Investments* 8 (1): 42–64.

Masters, M. W., and A. K. White. 2008. "The Accidental Hunt Brothers: How Institutional Investors Are Driving Up Food and Energy Prices." http://accidentalhuntbrothers.com/.

Pirrong, C. 2015. "Risk Management by Commodity Trading Firms: The Case of Trafigura." *Journal of Applied Corporate Finance* 27:19–26.

Plante, M., and M. Yücel. 2011. "Did Speculation Drive Oil Prices? Market Fundamentals Suggest Otherwise." *Federal Reserve Bank of Dallas Economic Letter* 6, no. 11, October.

Till, H. 2006. "Structural Sources of Return and Risk in Commodity Futures Investments." Working paper, EDHEC Risk and Asset Management Research Centre.

———. 2008. "Case Studies and Risk Management Lessons in Commodity Derivatives Trading." In *Risk Management in Commodity Markets: From Shipping to Agriculturals and Energy*, edited by H. Geman, 255–91. Chichester, UK: John Wiley & Sons.

———. 2009. "Has There Been Excessive Speculation in the US Oil Futures Markets?" Working paper, EDHEC Risk and Asset Management Research Centre.

Till, H., and J. Eagleeye. 2005. "Commodities: Active Strategies for Enhanced Return." *Journal of Wealth Management* 8 (2): 42–61. Also in *The Handbook of Inflation Hedging Investments*, edited by R. Greer, 127–57. New York: McGraw-Hill.

Working, H. 1949. "The Theory of the Price of Storage." *American Economic Review* 39:1254–62.

Zaremba, A. 2015. "Is Financialization Killing Commodity Investments?" *Journal of Alternative Investments* 18 (1): 66–91.

CHAPTER 23

Allocation to Commodities

This chapter provides a brief outline of the roles of commodity investments, and highlights some of the research that has helped foster interest in commodities as an investment. Many empirical results have shown the following:

- During certain periods, commodity investments provide both return enhancement and diversification benefits (as well as a hedge) for inflation risk, business cycle risk, and event risk.
- An allocation to commodities can improve the efficient frontier for a portfolio consisting of traditional as well as alternative investments.
- Some commodities have a low, sometimes even negative, correlation with one another, indicating that diversified portfolios of commodities can eliminate most of the idiosyncratic risks of individual commodities.

23.1 FIVE BENEFICIAL CHARACTERISTICS OF ALLOCATIONS TO COMMODITY FUTURES

The appeal of commodity futures in asset allocation has traditionally been based largely on five characteristics:

1. Low correlation with stocks and bonds, improving the risk-return profiles of portfolios
2. Inflation, business cycle, and event risk hedging abilities
3. Mean reversion and diversification return, leading to improved performance through rebalancing
4. Positive risk premium and roll return earned by investors in commodity futures
5. Return distributions that may be positively skewed

The next five sections provide detail on each of these five beneficial characteristics.

23.1.1 Low Correlation with Stocks and Bonds

It was not until the 1970s that academic studies highlighting the positive role of commodities in institutional and family office portfolios appeared.

23.1.1.1 The Effect of Full Collateralization on Perceived Commodity Risk In a seminal article published in the *Journal of Portfolio Management*, Greer (1978) tackled the issue of perceived risk in commodity futures. He showed that risk in a commodity futures position could be lowered significantly through full collateralization. Using a data set from 1960 to 1974, Greer calculated the returns of an unleveraged, fully collateralized basket of commodity futures contracts, including collateral returns. Greer's article also pioneered efforts to demonstrate the benefit of diversifying an equities-only portfolio with the addition of commodity futures, showing that a rebalanced portfolio of stocks and commodities provided a steadier and higher rate of return than a stock-only portfolio. Bodie and Rosansky (1980) echoed Greer's findings, showing that an equally weighted basket of commodity futures (constructed using 1950–76 data) provided valuable diversification benefits for an equity portfolio. Since then, a number of studies have demonstrated the benefits of adding commodities to portfolios of stocks and bonds.[1]

23.1.1.2 Liabilities, Futures, and Diversification with Commodities Commodities are particularly beneficial when employed by institutional investors, such as pension funds, endowments, foundations, and insurance companies. Payoffs from these funds are typically used to support liabilities that may grow in response to inflation and increases in commodity prices. Commodities could assist in an institution's pursuit of matching future asset returns against expected liabilities. A meaningful optimal allocation to commodities, based on their diversification and risk-reduction potentials, should be beneficial to such investors.[2]

While the diversification benefits of commodities can be realized using either spot or futures investments, the levels of return tend to be lower when spot contracts are used. The primary reason is that spot investments provide convenience yield, which is not a primary concern of portfolio investors. On the other hand, futures-based investments may be able to capture the convenience yield in terms of a higher rate of return. In addition, investments that provide indirect exposures to commodities (e.g., equities of commodity-producing firms) do not provide the same degree of diversification benefits as direct exposures to them.[3]

Commodities tend to perform better than certain alternative investments in providing diversification benefits for traditional portfolios. In particular, commodities provide more effective diversification than most institutional real estate investments. In addition, commodity futures have the added benefit of high liquidity, which many alternative investments may lack.[4]

23.1.1.3 The Evidence on Commodities as a Diversifier The diversification potential of commodity investment was further explored by Gorton and Rouwenhorst (2006) and later updated in Bhardwaj, Gorton, and Rouwenhorst (2015), who found that for time horizons longer than one month, an equally weighted commodity futures portfolio was negatively correlated with the return on both the S&P 500 and long-term bonds. The hypothesis that the correlation of commodity futures with stocks is zero at short horizons cannot be rejected, suggesting that commodity futures are effective in reducing the risk of equity and bond portfolios. The negative correlation of commodity futures with stocks and bonds tends to become more negative as the holding period increases; the observed negative correlations were found to be most pronounced at five-year holding periods. This pattern suggests that the diversification benefits of commodity futures may be best realized by investors with long horizons.

However, as the following exhibits show, these results are somewhat time dependent and are also a function of the commodity index selected and the weights assigned to each commodity.

Exhibits 23.1 and 23.2 report results that are different from those reported by Bhardwaj, Gorton, and Rouwenhorst (2015) and other older studies of the statistical properties of commodities. There are two primary reasons for this. First, the type of commodity index and the way it is constructed has a major impact on its behavior. The previous exhibits used the S&P GSCI, which, unlike the indices used in Bhardwaj, Gorton, and Rouwenhorst (2015), is not equally weighted. Second, it can be seen

EXHIBIT 23.1 Investment Performance of Commodity and Other Indices, Monthly Data, 1996–2015

					Correlations			
	Annualized Return	Annualized Standard Deviation	Skewness	Kurtosis	USD	Commodity Index	Stocks	Bonds
1996–2000								
USD[1]	5.3%	7.2%	−0.15	0.06	1.00	−0.11	−0.02	−0.84
Commodity index[2]	8.9%	20.4%	0.12	−0.06	−0.11	1.00	0.03	0.15
Stocks[3]	10.2%	14.4%	−0.95	1.88	−0.02	0.03	1.00	0.08
Bonds[4]	3.6%	5.4%	0.42	0.68	−0.84	0.15	0.08	1.00
2001–5								
USD	−3.6%	7.7%	−0.22	−0.17	1.00	−0.16	−0.14	−0.86
Commodity index	9.4%	22.1%	−0.21	−0.23	−0.16	1.00	0.04	0.10
Stocks	0.6%	14.8%	−0.58	0.34	−0.14	0.04	1.00	−0.07
Bonds	7.1%	5.9%	0.12	0.05	−0.86	0.10	−0.07	1.00
2006–10								
USD	−2.8%	9.7%	0.30	0.74	1.00	−0.59	−0.66	−0.83
Commodity index	−5.9%	28.9%	−1.12	2.87	−0.59	1.00	0.62	0.26
Stocks	0.4%	19.9%	−1.11	2.29	−0.66	0.62	1.00	0.40
Bonds	6.7%	6.8%	−0.19	0.64	−0.83	0.26	0.40	1.00
2011–15								
USD	4.5%	7.3%	0.42	0.09	1.00	−0.67	−0.55	−0.62
Commodity Index	−16.5%	19.9%	−0.52	0.10	−0.67	1.00	0.53	0.32
Stocks	5.2%	12.9%	−0.40	0.84	−0.55	0.53	1.00	0.32
Bonds	1.2%	3.9%	−0.43	0.48	−0.62	0.32	0.32	1.00

[1]U.S. Dollar Index spot rate.
[2]S&P GSCI Total Return CME.
[3]MSCI World USD.
[4]Barclays Capital Bond Composite Global.
Source: Bloomberg and authors' calculations.

EXHIBIT 23.2 Correlation Coefficients of Commodities with Traditional Assets, 1996–2015

1996–2015	USD Index	Global Bonds Index	Global Stocks—DEV Index	Global Stocks—EM Index	Comm. Index	Energy Index	Industrial Metals Index	Precious Metals Index	Agriculture Index
USD[1]	1.00	–0.80	–0.38	–0.33	–0.41	–0.35	–0.41	–0.42	–0.35
Global bonds[2]	–0.80	1.00	0.19	0.15	0.21	0.17	0.19	0.41	0.26
Global stocks—DEV[3]	–0.38	0.19	1.00	0.83	0.35	0.28	0.53	0.17	0.32
Global stocks—EM[4]	–0.33	0.15	0.83	1.00	0.40	0.33	0.57	0.31	0.33
Commodities[5]	–0.41	0.21	0.35	0.40	1.00	0.98	0.51	0.33	0.40
Energy	–0.35	0.17	0.28	0.33	0.98	1.00	0.40	0.25	0.24
Industrial metals	–0.41	0.19	0.53	0.57	0.51	0.40	1.00	0.36	0.34
Precious metals	–0.42	0.41	0.17	0.31	0.33	0.25	0.36	1.00	0.29
Agriculture	–0.35	0.26	0.32	0.33	0.40	0.24	0.34	0.29	1.00

[1]US Dollar Index spot rate.
[2]Barclays Capital Bond Composite Global.
[3]MSCI World USD.
[4]MSCI Emerging Markets USD.
[5]S&P GSCI Total Return CME.
Source: Bloomberg and authors' calculations.

from Exhibit 23.1 that the results are highly time dependent. In particular, in recent years, the correlation between the S&P GSCI and global stocks has increased.

While evidence exists of the historical diversification benefits of commodities, there is some evidence that these benefits might be much smaller going forward. Evidence suggests that correlation between commodities and other asset classes has been increasing in the past decade. The primary reason for this is the financialization of the commodity markets. In a sense, commodities may become the victims of their own success. As more research made the case for including commodities in traditional portfolios, more money flowed into this asset class. The result is that as an asset class, commodities have become more integrated with the rest of the investment universe; therefore, they are more affected by factors that influence the broader investment universe (see Chapter 22 of this book). Also, measuring the diversification benefits of commodities is not easy, because as an asset class, commodities are particularly heterogeneous (see Exhibit 23.2). Thus, the choice of commodity index, or the related choice of individual commodities, will have a significant impact on a portfolio's performance.[5]

23.1.2 Inflation, Business Cycle, and Event Risk Hedging

Commodities have traditionally been considered to be effective hedges for inflation, particularly for unexpected inflation. Earlier studies showed that equally weighted portfolios of commodity futures were effective inflation hedges. The benefits of hedging against unexpected increases in inflation are particularly useful for bond portfolios.[6]

23.1.2.1 Commodities, Inflation Hedging, and Investment Horizon Hedging benefits of commodity futures tend to increase as the investment horizon increases. Because commodity futures returns are more volatile than inflation rates, long-term correlations better capture the inflation-hedging properties of commodity investments. While stocks are negatively correlated with inflation over long horizons, and bonds are negatively correlated over shorter horizons, the correlation of commodity futures with inflation is positive at all horizons and statistically significant at the longer horizons. Exhibit 23.3 displays the correlation between an equally weighted portfolio of commodity futures and the U.S. inflation rate.

Whereas commodity indices have traditionally been considered effective hedges for inflation, individual commodities vary in their degree of inflation hedging. For

EXHIBIT 23.3 Correlations between Inflation and Commodities, Stocks, and Bonds, August 1959–December 2014

	Stocks	Bonds	Commodity Futures
Monthly	−0.08	−0.18	0.09
Quarterly	−0.08	−0.23	0.24
One year	−0.13	−0.29	0.33
Five years	−0.10	−0.20	0.47

Source: Bhardwaj, Gorton, and Rouwenhorst (2015).

instance, industrial and precious metals as well as energy products tend to provide better hedges against unexpected inflation in comparison to other commodities. In addition, different commodities provide a different degree of protection against unexpected inflation depending on the level of inflation. Historically, energy products have provided strong protection against high unexpected inflation. Thus, commodities provide valuable inflation-hedging properties, particularly in periods of increasing inflation. However, not all commodities have direct links with measured inflation, and some do not have a major economic impact on prices of goods and services that are used to measure inflation.[7]

23.1.2.2 Commodities as Diversifiers to Economic Cycles In addition to providing hedging against unexpected changes in inflation, commodity indices may provide exposure to long-term growth in global gross domestic product (GDP) and demand. As indicated in Chapter 22, commodities go through periods of significant positive returns known as supercycles, which are primarily driven by industrialization and urbanization of emerging economies.

Different commodities tend to display different correlations with the business cycle. Compared to stocks and bonds, commodities can provide more predictable returns as the economy moves through various stages of the business cycle. For instance, agricultural commodities provide a natural hedge against the business cycle, as—unlike stocks and industrial commodities—they are less affected by it. Therefore, they can reduce the sensitivity of a portfolio to changes in performance of the economy. On the other hand, industrial metals display a different relationship with the business cycle, as they are directly affected by changes in the level of economic activities. Consequently, commodities are unique in that they are effective diversifiers of systematic (theoretically nondiversifiable) risk.

23.1.2.3 Commodity Performance in the Four Major Business Cycle Phases Equities and commodities tend to perform similarly, on average, through an expansion and a recession. Interestingly, however, when one considers the phases of recessions and expansions, the results are markedly different. In the early phase of a recession, stocks and bonds exhibit low or negative returns, while commodity returns are positive. The relationship is the opposite in the late phase of a recession. In fact, commodities tend to outperform stocks and bonds in late expansion and early recession phases.[8] For example, assuming all else is equal, the following active strategy using stocks, bonds, and commodities would take advantage of the behavior of different asset classes during the four major business cycle phases (see Exhibit 23.4):

- Early expansion: long stocks, short commodities
- Late expansion: long commodities, short bonds
- Early recession: long commodities, short stocks
- Late recession: long bonds, short commodities

Another benefit of commodity investment, from an asset allocation perspective, is the ability of such investments to hedge event risk. Whereas non-commodity-related equity prices tend to be impacted negatively by natural disasters and political or economic distress, some commodity prices generally react positively to such occurrences. This characteristic of commodity returns is particularly beneficial, since it results in further reduced correlations between stocks and commodities in periods of

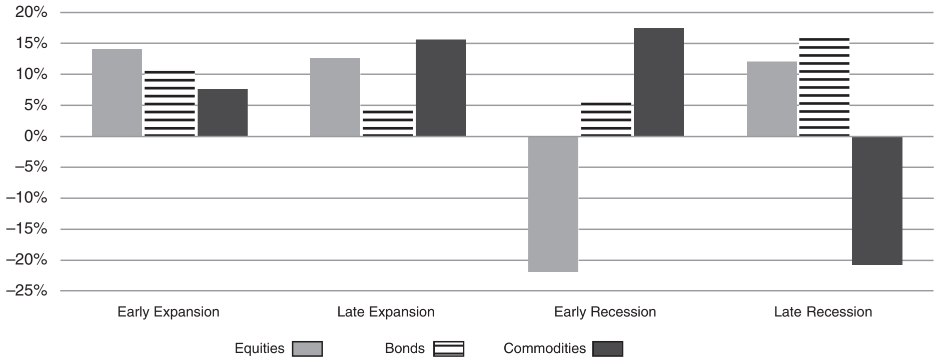


EXHIBIT 23.4 Average Returns by Stage of the Business Cycle, 1972– 2012
Sources: Dow Jones Indices (2012) and Bloomberg.

market stress caused by supply shocks. In addition to offering protection from major supply-driven market stresses, commodities can provide protection from more narrowly focused events as well as from unanticipated macro trends (or, likewise, exposure to any differentiated expectations related to macro trends). For example, a strike at a large zinc mine may decrease the price of the firm's equity while simultaneously putting upward pressure on zinc prices. In addition, long/short positions in commodities can provide a hedge against unanticipated global growth (or weakness). Recently, as the emerging economies of countries like China, India, Indonesia, South Africa, and Mexico expanded at a rapid pace, their need for raw materials, energy, and agricultural products put upward pressure on the global prices of certain commodities.[9]

23.1.3 Mean Reversion and Diversification Return

Commodities have long been considered to exhibit mean reversion in their prices. This section discusses the potential to generate diversification return in a market with mean-reverting prices.

23.1.3.1 Benefits of Mean Reversion in Commodity Investing Mean reversion can be a great advantage in an asset allocation framework. To the extent that assets exhibit this property, they may provide enhanced return from rebalancing. While the degree of mean reversion varies by commodity, the available evidence suggests that mean reversion at different time horizons is a common feature of commodities. It is often said that the cure for high prices is high prices. This means that as commodity prices rise because of supply or demand shocks, marginal producers that could not compete at previously low prices begin to supply the market, eventually reducing the price. By the same token, marginal users will leave the market, reducing the demand and forcing the price lower. Similarly, a decline in prices will eventually force marginal producers out of business, reducing the supply and thus increasing the price.[10]

23.1.3.2 Benefiting from Mean Reversion through Portfolio Rebalancing One approach to harvesting the benefits of mean reversion is through rebalancing (an action that brings portfolio allocations back into line with target allocations). The enhanced average or expected geometric mean return from rebalancing (or other volatility reduction) is often termed **diversification return**. Diversification return is not unique to commodities. The term was first coined by Booth and Fama (1992), and has been studied for equities, bonds, emerging markets, and multi-asset-class portfolios. Commodities seem particularly well suited for obtaining this source of benefit for two reasons:

1. Diversification returns are highest when the individual assets in a portfolio are highly volatile and the correlation among those assets is low.
2. Most importantly, commodity prices tend to display mean reversion.

Commodities have low correlations with one another and so can offer uncorrelated investment opportunities across various commodity markets. In particular, the energy sector does not have high positive correlation with other sectors, because higher energy prices, in particular, can weigh on economic growth and depress demand for other commodities. This correlation pattern can potentially generate diversification returns while lowering the risk of a diversified commodity portfolio.

The volatility of a portfolio will be reduced if the assets (e.g., commodities) included in the portfolio are not highly correlated with each other. Further, this increases the expected values of the most likely outcomes (not the expected future value of the entire distribution) from investing in commodities.

23.1.3.3 Volatility Reduction Enhances Geometric Mean Returns, Not Expected Values It is important to note that the increase in geometric mean return of the portfolio that results from lowering the portfolio's return volatility has no impact on the expected value of the portfolio. As discussed in Chapter 22, although reducing the volatility of a portfolio increases its compounded rate of return,[11] it does not affect the expected future value of the portfolio.[12] This means that in the absence of mean reversion, rebalancing will have no effect on the portfolio's future value. However, even in the case when volatility is lowered, the most likely outcomes tend to be higher in value. On the other hand, when there is mean reversion in prices, rebalancing may increase the expected future value of the portfolio as well. Although rebalancing's only impact is to increase the likelihood that the portfolio will make money, the lower volatility means that the best outcomes will not be as rewarding as when the volatility is high.

Diversification return is related to the fact that rebalancing reduces the portfolio allocation to commodities whose weights have increased since the beginning of the period, while increasing the allocation to commodities that have declined in relative value. This return to contrarian trading benefits when prices exhibit mean reversion (rather than momentum). Frequent rebalancing to initial allocations improves returns if asset values exhibit mean reversion and detracts if there is significant price momentum.

23.1.4 Positive Risk Premium

In his seminal article, Greer (1978) showed that an unleveraged, collateralized portfolio of commodity futures provided higher returns and a lower maximum drawdown than an index of equities. Subsequent studies supported Greer's findings, showing that an equally weighted portfolio of collateralized commodity futures produced equity-like returns.[13]

Following these early studies, a steady evolution of research has indicated that commodity index returns and volatility are similar to those of equities.[14] These studies found that a portfolio of commodity futures had roughly the same return and Sharpe ratio as U.S. equities. However, in looking at the equities of firms that produce commodities, the firms' stocks were more highly correlated with the stock market than with the corresponding commodity futures. This is reinforced by the fact that an equity index composed of commodity producers has underperformed a futures-based commodity index. Recent studies that used the S&P GSCI from 1970 to 2009 have found that commodities provided similar return and risk performance to U.S. and international equities. In particular, Bhardwaj, Gorton, and Rouwenhorst (2015) found that fully collateralized portfolios of commodity futures provided returns and volatilities very similar to those of stocks. The average historical risk premium of commodity futures was about equal to that of stocks at 5% per annum, which is almost twice that of bonds.

As discussed in Chapter 22, commodities that are difficult to store (either because storage is impossible or expensive, or because it is more efficient to leave the commodity stored belowground rather than extracting it) are more likely to be in backwardation and to generate positive returns. In contrast, difficult-to-store commodities tend to exhibit high spot price volatility; low inventories can increase price volatility by exacerbating the impact of supply-and-demand shocks. Consistent with theory, futures contracts on difficult-to-store commodities (e.g., crude oil, gasoline, live cattle, live hogs, and soybean meal) have provided positive returns in the past. In particular, the return on a commodity index is proportional to the amount of time the commodity is in backwardation.

The positive return from commodities is not uniformly available during all periods, and, given the preceding discussion, it is evident that it varies among commodities. This indicates that a more active approach to commodity investment is likely to produce better risk-return profiles. In particular, a long position can be taken only in those commodities that display backwardation or report low inventory levels. Measures such as slope of the futures curve, inventories, and even momentum provide useful information about commodity futures risk premiums. For example, low levels of inventory may increase the hedging demand by users of commodities, creating risk premiums for speculators who are willing to take short positions in such markets.

Due to the statistical properties of commodities just discussed, most investors can justify a commodity allocation even if the average risk premium is low. More importantly, as shown by a number of studies, positive risk premiums provided by commodity investments can be significantly improved upon by strategically allocating across commodities that display momentum and backwardation.[15]

23.1.5 Positive Skewness

Equity and equity-related investments tend to display negative skewness. It is relatively easy to think of events that could cause large negative returns to these investments. Political risk, systemic risk, credit events, and so on produce large negative returns on equities.

Commodity prices are also affected by events. Supply shocks tend to be larger and less predictable than demand shocks. As a result, commodity prices are expected to display large positive returns more frequently than what is predicted by a normal distribution. Therefore, compared to a normal distribution, return distributions of commodities are expected to display positive skewness. However, it is important to note that positive skewness is present during specific periods, and the level of skewness is not always significantly different from zero. Exhibit 23.5 displays correlations among commodity indices for various subperiods.

Since investors tend to prefer positive skewness, this feature of commodity returns makes them somewhat attractive to investors. However, compared to the benefits discussed previously, this plays a rather small role in making commodities attractive. Also, commodities tend to have higher kurtosis than do traditional investments, and there is some evidence that investors prefer optimal portfolios with low kurtosis (smaller tail risks). This feature of the return distribution of commodities is strong and statistically significant and may offset the potential benefits from having positively skewed return distributions.[16]

23.2 COMMODITY INVESTMENT STRATEGIES

In contrast to traditional stock and bond markets, research has shown that commodities offer unique risk and return alternatives. These risk and return alternatives are due to the unique supply-and-demand conditions affecting physical spot markets as well as the changing carry and storage relationships that influence futures pricing. When one adds micro market impacts related to the unique trading processes that exist in commodity markets, including rollover and delivery impacts, it is evident that there may exist numerous potential opportunities to generate profits. Questions remain, however, as to what degree commodities can offer returns consistent with their underlying risk exposure, including whether various commodity investment strategies offer the potential for generating alpha. Most strategies involve trading either commodities directly or the companies whose profitability is predominantly driven by underlying commodity pricing. Commodity trading strategies can be separated into two broad groups: directional and relative value. **Directional strategies** take positions substantially exposed to systematic risk based on a forecast of market direction. Relative value strategies attempt to identify mispriced assets or securities and to hedge away some or all of the market exposure.

23.3 DIRECTIONAL STRATEGIES

Directional strategies are those that express a view by betting on a market's direction by taking either a long or a short position. While these strategies can also be implemented with traditional assets, such as the equities of commodity firms or commodity-based exchange-traded funds (ETFs), they more often use listed commodity derivatives (futures and options) as well as over-the-counter (OTC) derivatives, such as forward contracts, swaps, or swaptions. Some strategies also involve holding physical commodities, though this is much less common, primarily due to the high cash costs to purchase and store them.

EXHIBIT 23.5 Correlations among Commodities Subindices

GSCI Subindices	S&P GSCI	Energy	Industrial Metals	Precious Metals	Gold	Wheat	Corn	Soybeans	Cotton	Sugar	Coffee	Cocoa
1996–2000	0.12	0.05	0.12	1.31	1.57	0.24	−0.24	−0.08	0.22	−0.22	0.24	0.24
2001–5	−0.21	−0.07	0.23	−0.38	−0.21	0.38	−0.09	−0.35	0.28	−0.23	0.65	0.27
2006–10	−1.12	−0.91	−0.80	−0.80	−0.77	0.17	−0.41	−0.77	−0.11	−0.19	−0.19	−0.22
2011–15	−0.52	−0.51	−0.83	0.03	0.02	−0.31	0.45	−0.04	−0.61	0.39	1.26	−0.83

Source: Bloomberg and authors' calculations.

In the context of commodities, **fundamental directional strategies** implement allocations based on an analysis of the underlying supply-and-demand factors for commodities or commodity sectors. They are often based on macroeconomic factors such as economic growth, interest rate, and currency forecasts, or on specific factors driving a commodity's price, such as the number of cattle in feedlots or changes in weather conditions.

In the context of commodities, **quantitative directional strategies** use technical or quantitative models to identify overpriced and underpriced commodities based on spot price forecasts or mispriced futures term structures. Most of these quantitative strategies try to take advantage of special features of commodity futures markets, features that are not available in traditional equity or bond markets. Chapter 22 discussed the concepts of roll yield, convenience yield, and risk premium resulting from hedging pressure. All of these could be used as the basis for long and short quantitative commodity investment strategies.

Trading strategies based on convenience yield and potential roll return have these features. First, the potential roll yield is estimated using the price differential between front and second nearest contracts. Second, historical roll yield is used to obtain estimates of future roll yields. This basically assumes that there is a momentum or a trend in roll return. Third, depending on estimates of the potential yield, an equally weighted long/short portfolio is created. Long positions are established in commodities that are expected to have positive roll returns, and short positions are taken where negative roll returns are expected. Fourth, levels of inventory or other measures of scarcity are used to supplement the current information extracted from the shape of the yield curve. As discussed previously, lower inventory levels are associated with higher levels of backwardation and higher roll return. Finally, the positions are rebalanced after a set period, typically one month. This simple long/short strategy has significantly outperformed the long-only strategy in the past.[17]

Roll yield is one potential source of return from futures contracts. The other source of return is the risk premium embedded in futures prices. Normal backwardation, which was discussed in Chapter 22, claims that most hedgers are commodity producers and are net short commodity futures. Therefore, speculators who take long positions in futures contracts will earn a risk premium, because it is argued that futures prices are below expected future spot prices. Empirical evidence does not support this simple version of futures prices and risk premium. The primary reason is that hedgers may not always be net short commodity futures. In other words, if users of commodities are the main hedgers in futures markets, then they would want to take long positions in futures contracts, and their hedging pressure will raise the current futures price above the expected future spot price. In other words, when hedging pressure is coming from those who take long positions in futures contracts, the risk premium will be negative and will be earned by those speculators who are willing to take short positions in the futures contracts. Estimating the potential risk premium in commodity futures is a more difficult task. Researchers have shown that the hedging pressure by long hedgers is greatest when markets are most backwardated. Also, some exchanges report positions held by different groups of market participants. This information can be used to estimate the size and sign of the risk premium. Finally, traditional measures of riskiness, such as the capital asset pricing model (CAPM) and factor models, have been shown to be useful in predicting the risk premium.[18]

It is important to note that while spot commodity investments will benefit from a rise in commodity prices, investments in futures contracts will benefit to the degree that the increase in the spot price was not anticipated by futures prices. For example, if the futures curve is upward sloping, then the futures market is anticipating some increase in spot prices. An investor who takes a long position in, say, a six-month contract will benefit from a price rise in the spot price of the commodity over the next six months only if the spot price rises above the futures price.

Specific applications of these models and the development of related strategies are further discussed in Chapter 24.

23.4 RELATIVE VALUE STRATEGIES

Relative value strategies in commodity markets are best understood as businesses rather than as trading strategies. Relative value managers combine investment capital with expertise in a particular commodity sector to provide economic value added. They generally compete for profits with vertically integrated commodity firms, merchant banks, shipping companies, and trading firms.

Relative value strategies in commodities can be executed across three risk dimensions: (1) location, (2) correlation, and (3) time. The same commodity can have different prices at different locations in the world. The relative prices of two similar commodities can diverge from historical norms. The price of the same commodity can be different based on when the commodity is scheduled for delivery. For example, consider a spread trade that is long crude oil for delivery in October in the United Kingdom and short heating oil for delivery in December in the United States. This trade has all three risk dimensions: location, correlation, and time.

In contrast, relative value arbitrage strategies in equity and fixed-income markets are generally limited to two dimensions: correlation and time. A share of stock, irrespective of currency, sells at the same price everywhere in the world (or else allows for arbitrage profits), so there is no location dimension. Of course, this does not apply to equity derivatives. Commodity traders have more degrees of freedom when designing strategies.

23.5 COMMODITY FUTURES AND OPTIONS SPREADS

Commodity spreads are strategies that seek to take advantage of trading opportunities based on relative commodity prices that can be executed entirely in derivatives markets. These can involve futures contracts, forward contracts, OTC swaps, options, and swaptions.

23.5.1 Calendar Spreads

Perhaps the simplest strategy is a calendar spread, which involves taking opposing long/short positions in the futures market for delivery at different times in the future. These trades can be designed to provide liquidity or insurance against an unforeseen event, or to express a view.

23.5.1.1 Common Calendar Spreads For example, a typical calendar spread trade involves selling natural gas futures for delivery in March and taking a long position for delivery of natural gas in April. This spread could be profitable in a very mild winter but could lose money in a very cold winter. Utilities purchase more natural gas for delivery in March than they expect to need to ensure adequate supplies. If the winter is mild, they will sell the surplus gas inventories toward the end of the season, which will push down the price for delivery in March. Traders who hold this spread are writing a **synthetic weather derivative**, since it is a derivative position with returns that are substantially driven by weather conditions. They will earn a profit if the weather is mild but will take a loss if the weather is severe. Traders may also take the other side of this trade by buying natural gas for March delivery and selling it for delivery in April. If a natural gas shortage develops due to an exceptionally cold winter or supply disruptions, this trade will be profitable.

Other common calendar spreads are designed to provide liquidity. For example, in 2005 and 2006, it was common for traders to buy crude oil futures for delivery two months forward and hedge this position by selling crude oil for delivery three months forward. This trade was designed to take advantage of the growth in commodity index investments. Since most commodity index investments are designed to hold futures for near-term delivery, there was a surplus of these contracts on roll dates, when the indices were selling their short-maturity futures positions, and a corresponding shortage of second-deferred futures, when the indices were buying the second-deferred futures positions. Speculators responded by designing a calendar spread that offered liquidity to the indices at prices that could provide a reasonable expected profit to speculators. A study by Mou (2011) provides evidence that significant alpha could have been generated by exploiting the established roll pattern of the S&P GSCI. Mou tested two pre-roll strategies (five days and 10 days prior to the S&P GSCI roll) on 19 of the commodities included in the S&P GSCI (representing a weighting of about 93% of the S&P GSCI). Both pre-roll strategies generated significant abnormal returns over the period from 2000 to 2010, with the energy sector providing the best performance. Over the period from 1980 to 2000, however, a period in which there was little or no investment in the S&P GSCI, pre-roll strategies were not able to generate alpha. This result suggests that the alpha found for the pre-roll strategies from 2000 to 2010 was capturing price imbalances caused by the S&P GSCI roll process. Since 2006, however, the profitability of pre-roll strategies has been declining, to the point that these strategies may now have average excess returns of close to zero.

23.5.1.2 Bull Spreads and Bear Spreads Depending on their market views, traders can enter two types of calendar spreads: bull spreads and bear spreads. In a bull spread, the speculator is long the nearby (near-term) contract and short the distant (long-term) contract. In markets that are in backwardation, the investor is hoping for the spread to widen; in contango markets, the bull spread investor is hoping for the price difference to narrow. The losses of a bull spread investor tend to be limited because, in an efficient market, price differences cannot exceed carrying costs (adjusted for convenience yields). If at some point the differences do exceed carrying costs, arbitrageurs would drive prices toward a level reflecting fair carrying costs.

In a bear spread, a trader is long the distant (long-term) contract and short the nearby (near-term) contract. In markets in backwardation, the bear spread speculator

EXHIBIT 23.6 Profit and Loss Calculation

	July (1st Leg)	December (2nd Leg)	Total
March, open	Long: −\$44.37	Short: +\$50.78	
June, close	Short: +\$35.18	Long: −\$38.16	
Net per barrel	−\$9.19	\$12.62	\$3.43

wants the spread to narrow, whereas in contango markets, the bear spread speculator wants the price difference to widen. In theory, it is possible to see a rise in the price of the near-term contract without a similar rise in the price of the long-term contract (e.g., due to a workers' strike or weather). Therefore, in principle, the potential risk in a bear spread could be unlimited or at least very large.

23.5.1.3 Estimating the Profitability of Spread Trading Since spread trading is not an arbitrage strategy, the profit and loss (P&L) from a spread position can be calculated only after the spread is closed. Assume the following scenario (see Exhibit 23.6). In March, a spread trader observes a contango in the crude oil forward curve. Anticipating a flattening of the curve and a narrowing of the spread, the trader goes long three July light sweet crude oil futures contracts (traded on NYMEX) at \$44.37, simultaneously shorting three December light sweet crude oil futures at \$50.78. In April and May, an oversupply of crude in the world markets causes prices to slump across the board. At the beginning of June, the spreader closes out the July contract at \$35.18 and the December contract at \$38.16.

The profit for the combined position can now be determined. The first step, shown in Exhibit 23.6, is to compute the profit or loss per unit of the underlying commodity. The second step is to estimate the total profit or loss by multiplying the profit or loss per underlying commodity unit times the number of commodity units per contract and the number of contracts, as shown in Equation 23.1.

$$\begin{aligned}\text{Position P\&L} &= \text{P\&L}_{\text{barrel}} \times \text{Contract} \times \text{Position Size} \\ &= \$3.43 \times 1{,}000 \times 3 = \$10{,}290\end{aligned} \tag{23.1}$$

Since the size of the NYMEX light sweet crude oil contract is 1,000 barrels, and the size of the trader's spread position is three contracts, the total gain of the trader's position (ignoring trading and financing costs and collateral yield) is \$10,290.

APPLICATION 23.5.1.3

Wheat contracts typically trade in contracts for delivery of 5,000 bushels. Assume that May wheat contracts currently sell for \$5.50 per bushel, which is \$0.20 more than the futures price of wheat contracts with delivery in July, due to anticipated harvesting. A commodity trader with a sophisticated weather forecasting model anticipates highly abnormal weather that will drive the two futures prices to being equal. If the trader is correct, what will the profit be on a calendar spread position that is long 30 July contracts and short 30 May contracts?

It is not necessary to know absolute prices in calculating the profit or loss of a spread trade. In this example, the spread between the two dates is assumed to decline from $0.20 per bushel to $0.00 per bushel. The $0.20 per bushel profit is multiplied by 5,000 bushels times 30 contracts to calculate the total profit on the position of $30,000 (i.e., using Equation 23.1).

23.5.2 Processing Spreads

Processing spreads seek to take advantage of the relative price difference between a commodity and the products the commodity produces. For example, processing spread strategies can involve crude oil and its refined products. A common method of trading this processing spread involves buying crude oil (spot or futures) while selling a combination of products derived from crude oil (heating oil and gasoline). Another example of a commodity processing spread involves buying soybeans and selling a combination of soybean meal and soybean oil.

Processing spreads are frequently used by producers to lock in favorable margins. For example, a soybean processor may wish to buy the spread in the futures market (buy soybean futures, and sell soybean meal and soybean oil futures). However, there are no natural sellers of this spread. Instead, there are three separate natural participants on the other side of the spread:

1. Farmers who sell soybean futures
2. Livestock feed producers who buy soybean meal
3. Vegetable oil manufacturers who buy soybean oil

Speculators may provide liquidity by selling the spread to permit the processor to lock in a favorable margin and then either (1) immediately attempt to find buyers or sellers to unwind the individual components of their spread trade, or (2) keep the spread position and hope to unwind the entire transaction at a more favorable price.

There are two conditions that hold for all producers:

1. Producers take long futures positions to hedge against rising input prices.
2. Producers take short futures positions to hedge against falling output prices.

In other words, producers who choose to hedge are generally long the primary commodity in a nearer-term expiration month but also short its constituents in a later period.

It should be noted that futures exchanges set lower margins for bona fide hedging spreads, in which a producer goes long futures on the input and goes short futures on the output. On the flip side, a spread trader who goes long the output and short the input is subject to higher margin requirements despite providing hedgers with needed liquidity. Two examples of typical processing spreads follow.

Crack spreads are hedges typically used by oil refineries reflecting the cracking or breaking apart of large oil molecules in the process of refining crude oil into gasoline and heating oil. The typical producer hedge position would involve going long crude oil futures (to hedge future input purchases) and short gasoline and heating oil (to hedge against potential decreases in the price of the outputs, or distillates). The crack spread locks in the refiner's margin, which is the revenue earned for refining each barrel of crude oil.

It is a common practice to express the crack spread in terms of a ratio as X:Y:Z, where X represents the number of barrels of crude oil, Y represents the number of barrels of gasoline, and Z represents the number of barrels of heating oil, subject to the constraint that $X = Y + Z$. Typical crack spreads are 3:2:1, 5:3:2, and 2:1:1; as 3:2:1 is the most common multiple product crack spread, examples of U.S. crack spread benchmarks include Gulf Coast 3:2:1 and Chicago 3:2:1.

Crush spreads are hedges typically used by soybean processors, with its name derived from the physical crushing of soybeans into oil and meal. A typical crush spread would involve going long soybean futures (to ensure the processor against potential input price increases) and short soybean oil futures and soybean meal futures (to ensure against potential output price decreases). The analysis of crush spreads is very similar to the prior example presented for crack spreads.

23.5.3 Substitution Commodity Spreads

Substitution spreads are trades between commodities that can be substituted for one another in terms of either production or consumption.

23.5.3.1 Two Major Types of Substitutions in Commodities There are two types of substitutions:

1. A producer may use the same capital equipment to produce different products. Examples of producer substitution include corn and soybeans, because land suitable for growing soybeans is usually also suitable for growing corn. Oil refineries can also vary the mix of refined products. A refinery can be adjusted, within limits, to favor production of heating oil, jet fuel, or gasoline, depending on seasonal demand.
2. A consumer may be able to substitute one commodity for another based on their relative prices, whereby the consumer may have the flexibility to substitute with a less expensive commodity. As examples, utilities can use different fuels for electricity generation (e.g., natural gas can be substituted for oil-based fuels in the long run); cattle and hogs are also substitutes.

Substitution spread trades are generally riskier than processing spreads or calendar spreads, since they depend on historical correlations that may not persist or reappear in the future.

23.5.3.2 Determining Entry and Exit Points with a Substitution Test Statistic The general premise of substitution spreads is that the relative pricing relationship between easily substitutable commodities should be stable. If the price of one becomes too expensive, consumers will switch to the substitute. This results in a price drop for the original and a price rise for the substitute, forcing the ratio back to normal. To normalize ratios of contracts with different pricing specifications and contract sizes, one can study natural logs of the ratios of prices. The reason for this type of normalization is that price distributions can be approximated by a lognormal distribution, and therefore the distribution of the resulting ratio can be approximated by a normal distribution. This facilitates the testing of the difference to see if it is statistically significant. For example, one might look at the natural log of a series of ratios of heating oil to natural gas prices, such as:

$$\text{Substitute Test Statistic}_t = \ln\left[\frac{\text{Close Price(Heating Oil)}_t}{\text{Close Price(Natural Gas)}_t}\right] \tag{23.2}$$

For the purposes of illustration, let's assume that the nearest-maturity NYMEX heating oil futures contract (HO) closed at \$2.5620 per gallon yesterday, while the same maturity NYMEX Henry Hub natural gas futures contract (NG) closed at \$8.112 per 10,000 million British thermal units (MMBtu) yesterday. The HO/NG test statistic for yesterday would be estimated as follows:

$$\text{Substitute Test Statistic}_{\text{yesterday}} = \ln\left(\frac{\text{HO}_{\text{yesterday}}}{\text{NG}_{\text{yesterday}}}\right) = \ln\left(\frac{2.5620}{8.112}\right) = -1.15$$

To determine whether the spread has experienced a change significant enough to warrant a spread trade, a measure of stability is required. One such measure at traders' disposal is the difference from a 100-day moving average, calculated as follows:

$$\text{100-Day Statistic} = \frac{\text{Substitution Statistics}_t - \left(\text{100-Day MA of Substitution Statistics}\right)}{\text{100-Day Standard Deviation of Substitution Statistic}} \tag{23.3}$$

The critical values of the statistic that would trigger entering or exiting a spread position are determined statistically (optimizing over historical series of logs of price ratios of related securities).

Entering a spread long means going long the product in the numerator of the price ratio and shorting the product in the denominator. The assumption of the long spread position is that the product in the denominator has become too expensive relative to the product in the numerator. Shorting the spread means the opposite: shorting the product in the numerator and going long the product in the denominator. The assumption of the short spread position is that the product in the numerator has become too expensive relative to the product in the denominator.

23.5.3.3 Example of a Substitution Spread Trade For example, assume that the critical value of a substitute trade statistic for entry is 2.75, and the critical value for exit is 0. A long entry into the spread would be triggered if the 100-day statistic fell below –2.75 (i.e., execute pair trade: long heating oil, short natural gas), whereas a short position in the spread would be established if the 100-day statistic rose above

2.75 (i.e., execute pair trade: short heating oil, long natural gas). Long spreads are exited when the 100-day statistic rises above zero, and short spreads are exited when the 100-day statistic falls below zero.

In practice, though, the price reversion of a substitution spread between natural gas and heating oil would take place in the long term. Consider a common discussion circa 2012, when U.S. natural gas prices were reaching multiyear lows below \$2 per MMBtu, while diesel fuel was above \$4 per gallon at retail fueling stations. There were calls for shipping companies owning large fleets of trucks to convert from using diesel fuel to using natural gas. Despite the substantial difference in fuel costs at that point in time, switching costs were a significant consideration. In addition to switching truck engines from being powered by diesel fuel to being powered by natural gas, the infrastructure of natural gas fueling stations would have to be developed over highways carrying large portions of trucking traffic. Only after these longer-term investments were made could the demand for natural gas increase and the demand for heating oil decrease to the point that the Btu-equivalent price of the two fuels would start to meaningfully converge based on their general relationship as substitute goods.

23.5.3.4 Quality and Location Spreads **Quality spreads** are similar to substitution spreads, except that the spread is across different grades of the same commodity. A common quality spread executable in futures markets involves spring wheat and hard red winter wheat. Most other quality spreads involve OTC transactions. For example, there is a liquid OTC market in jet fuel, which is very similar to diesel fuel/heating oil. Similarly, there are many grades of coffee that are traded OTC but only a few grades that are listed on futures exchanges.

Quality spread traders provide liquidity to producers in the OTC market and then hedge by using other OTC markets or listed futures contracts. Most quality spreads have typically had relatively low price risk. Historically, there has been little chance that the spread between jet fuel and heating oil would undergo a major structural change, because refineries have easily switched production between heating oil and jet fuel when a demand imbalance takes place. Rather than list separate futures contracts for diesel fuel, jet fuel, and heating oil, liquidity in the futures markets is combined into a single commodity (heating oil), and users of the other two commodities use quality spreads to provide an effective hedge.

Location spreads are trades that involve the same commodity but different delivery and storage locations. A common location spread involves Brent crude oil, delivered in the United Kingdom, and West Texas Intermediate (WTI) crude oil, delivered in the United States. Location spreads are primarily traded using OTC derivatives, though some location spreads, such as the Brent/WTI spread, can be executed using listed futures contracts (Pringle and Fernandes 2007).

Some location spreads have an arbitrage component. For example, if the Brent/WTI crude futures trade is executed with a one-month lag, then it is possible to take delivery in the United Kingdom, incur transportation and storage costs, and then make delivery in the United States a month later. However, if the location spread is made without a lag, then the trade is a **correlation trade**, which is a trade with an outcome that is driven by the statistical correlation between two values, in this case the values of two commodities that differ by location.

23.5.4 Intramarket Relative Value Strategies

Many of the strategies used by active commodity managers and hedge funds combine the trading of commodity derivatives with the trading of the underlying commodities in the physical or spot markets. These strategies are designed to either implement an observed arbitrage pricing opportunity or use commodity futures markets the same way that commercial users do: as hedging vehicles to minimize the exposure of a particular transaction to unexpected market risk.

Storage strategies seek to profit from changes in the benefits and costs of commodity storage and often use leased storage facilities to hold physical commodities for delivery at a later date, when the return on storing a commodity exceeds its costs. These strategies are more complex than futures-based strategies and are both labor and capital intensive. Storage strategies are typically hedged transactions, involving a simultaneous purchase of the physical asset and sale of the commodity in the futures or OTC forward market. The strategy can also be an unhedged or a directional trade in anticipation of an increase in the commodity price, though this is less common. A storage strategy is equivalent to a calendar spread in that the transaction involves holding the same commodity over time.

Transportation strategies use spot commodity markets to execute location trades by moving commodities when the benefits of price differentials exceed transportation costs. The strategy involves leased transportation services, such as tankers, bulk shipping, or pipelines, to physically move a commodity from a location where the commodity is in surplus to a location where there is a shortage.

Certain transportation and storage strategies carry risks that futures-based strategies do not. For example, market participants using OTC contracts must be willing to assume the attendant credit risk of counterparties used in these transactions (Till and Eagleeye 2005). The participants must also be prepared to bear the risks associated with storing and transporting potentially hazardous commodities, including the potential headline and reputational risks. Market participants must also make sure that these strategies do not give the appearance of physical market manipulation. Till (2008) describes a Commodity Futures Trading Commission and U.S. Department of Justice action against a major international oil company, in which the company was fined $303 million for attempting to manipulate one U.S. delivery location's physical propane market. The firm's positions were initially entered into through the forward OTC market. This case was particularly memorable since the firm had actually failed in its attempted manipulation, losing at least $10 million in trying to carry out this cornering of the market.

23.6 CAPITAL STRUCTURE AND COMMODITY-BASED CORPORATIONS

Commodity-based corporations are typically valued as the sum of the firm's commodity rights and its enterprise value. **Commodity rights** reflect the current value of untapped commodity assets, such as oil reserves. The **enterprise value** is the residual value of corporate assets, equal to common equity plus preferred stock plus debt (less cash and other non-operating assets). Thus, equity and debt of commodity-based firms tend to act as hybrid instruments, combining the risk exposures of the corresponding asset class with a long or short exposure to commodity risk. The commodity betas of these securities are directly affected by a variety of firm characteristics.

23.6.1 Commodity Risk Management Strategies of Commodity Producers

Hedging policy has the most direct impact on commodity exposure. Hedging policy can vary significantly both across industries and within a particular industry. While gold hedging policies have generally moved away from hedging for the last decade due to elevated gold prices, traditionally some gold producers such as Barrick Gold would strive to hedge all of their gold price exposure (although others would hedge none). Adding a level of complexity to the analysis, Stulz (1996) provides evidence that gold producers alter their hedging policy based on their outlook for gold prices, a practice that he terms selective hedging. **Selective hedging** is the attempt to add value by market-timing the degree to which risk is hedged. This introduces a significant nonlinearity into the commodity exposure of firms following such practices.

In addition to a firm's direct hedging policy, its commodity exposure is further influenced by operational hedging. **Operational hedging** is the attempt by a firm to add value by altering its physical activities in reaction to commodity price changes to mitigate the impact on profitability. The adjustments may be in the form of timing, location, input quality, and so on, and often involve embedded real options. While direct hedging reduces risk, operational hedging may reduce risk and add value.

Operational diversification occurs across commodities, grades of commodity, and location to further reduce aggregate commodity price risk exposure. Typically, the equity and debt of commodity-based firms tend to be much less of a pure play than direct commodity investment or commodity futures investment.[19] Commodity-based firms tend to be diversified across commodities (horizontally integrated) or across business and risk exposure (vertically integrated). In the extreme, a vertically integrated commodity firm may have perfectly balanced long and short exposures to commodity prices across its operations. Highly integrated firms tend to have much less exposure to commodity prices and lower earnings volatility.

23.6.2 Commodity Price Risk of the Securities of Commodity Firms

Upstream commodity producers are direct producers (extractors or growers) of commodities, while **downstream producers** process or refine the output of the upstream producers into a marketable product. Generally speaking, upstream producers tend to have the highest exposure to commodity prices.

As financial leverage and cost structure have a compounded impact on the default and market risk of a firm, they also impact the sensitivity of firm's securities to commodity prices. The earnings of commodity-based firms with tight profit margins tend to be very sensitive to movements in commodity prices relative to other firms in the same industry. Thus, their equity and debt prices will fluctuate more with movements in commodity prices. High financial leverage results in a similar increased sensitivity.

23.6.3 Commodity-Based Equity and Debt Investment Strategies

For those seeking commodity exposure, commodity-based equity and debt provide an alternative to commodity index-based investment products. Such positions can be executed directly or combined with a short equity position (e.g., short stock index, long index puts) to hedge the market exposure and maintain the commodity exposure.

In other instances, the investor may desire the equity or debt exposure without the commodity exposure. For example, a long position in the equity of an airline may be hedged by buying physical jet fuel as a hedge against the impact of an increase in the price of jet fuel on the profitability of the airline.[20] This hedge effectively eliminates the commodity exposure of the stock, leaving the equity market exposure intact. If the firm is a commodity producer, the investor would typically hedge the commodity exposure by selling futures contracts or buying put options on commodity futures.

If an analyst determines that the equity value of a commodity producer is overpriced (unrelated to the future direction of commodity prices), a common strategy is to sell the firm's stock short and buy commodity futures or call options to hedge out exposure to the commodity rights. Conversely, if the stock is perceived to be underpriced (again, separately from changing commodity prices), the strategy involves buying the equity and selling futures or buying puts to hedge the commodity rights.

Debt strategies use commodity futures and options to hedge the commodity price component of the default risk of commodity producers and consumers. As previously discussed, the default risk associated with commodity prices can vary significantly across firms, largely driven by cost structure and financial leverage. Because default risk is highly nonlinear, commodity options are commonly used for this type of hedge. For example, a lender to a copper producer might purchase put options on copper futures to hedge this exposure.[21] An institutional investor owning an airline's bond might decide to purchase OTC call options on jet fuel or calls on heating oil futures to hedge the negative effect that higher fuel costs will have on cash flow (if the airline is not adequately hedged).

23.7 CONCLUSION

This chapter has examined the case for commodities as an asset class that should be considered by institutional investors, highlighting the benefits of this type of allocation. It discussed the role of commodities as hedges against unexpected inflation, as well as the diversification benefits that they can provide at different stages of the business cycle. It was pointed out that the potential diversification contributions of commodities to traditional portfolios may be diminishing over time due to financialization of commodity markets. Finally, the chapter examined various strategies for investing in commodities and discussed the importance of investing in products that include a diversified portfolio of commodities. While compounded returns to individual commodities have been low historically, compounded returns to diversified portfolios of commodities have been positive and similar to those from equities.

NOTES

1. Satyanarayan and Varangis (1994); Idzorek (2006). For a recent study, see Bessler and Wolff (2015); Benham, Walsh, and Obregon (2015).
2. Beenen (2005).
3. Froot (1995).
4. Ankrim and Hensel (1993); Satyanarayan and Varangis (1994).
5. See Tang and Xiong (2012); Erb and Harvey (2006); Gospodinov and Ng (2013).
6. See Bodie and Rosansky (1980) and Froot (1995).
7. Fama and French (1988); Bjornson and Carter (1997).
8. Gorton and Rouwenhorst (2006); Dow Jones Indices (2012); Benham et al. (2015).
9. Chong and Miffre (2008); Burkart (2006).
10. Bessembinder et al. (1995).
11. See Erb and Harvey (2006); Till and Eagleeye (2005); Willenbrock (2011).
12. See Chambers and Zdanowicz (2014).
13. Bodie and Rosansky (1980).
14. Greer (2000); Gorton and Rouwenhorst (2006); Bhardwaj, Gorton, and Rouwenhorst (2015).
15. Miffre and Rallis (2007); De Groot, Karstanje, and Zhou (2014).
16. Kat and Oomen (2006); Jensen and Mercer (2011); Bhardwaj, Gorton, and Rouwenhorst (2015).
17. See Blitz and De Groot (2014); Miffre (2015).
18. Miffre (2015); Basu and Miffre (2013).
19. It is worth noting that the commodity price component of a stock price is more forward looking than futures prices since the stock price is a reflection of the present value of cash flows with an infinite time horizon whereas the time horizon of a futures price is the expiration date of the contract. This suggests that commodity-based equities may provide higher leverage to commodity price movements than commodity futures do.
20. This hedge would typically be executed with a long position in a futures contract. Since futures contracts are not available on jet fuel, a long position in heating oil futures would be used.
21. There are many alternatives to puts, including credit default swaps (CDSs) or equity shorts on a basket of copper miners (while also hedging exposure to higher equity prices).

REFERENCES

Ankrim, E., and C. Hensel. 1993. “Commodities in Asset Allocation: A Real-Asset Alternative to Real Estate.” *Financial Analysts Journal* 49 (3): 20–29.

Basu, D., and J. Miffre. 2013. “Capturing the Risk Premium of Commodity Futures: The Role of Hedging Pressure.” *Journal of Banking and Finance* 37 (3): 2652–64.

Beenen, J. 2005. “Commodity Investing: A Pension Fund Perspective.” *Futures Industry*, September/October, 18–22.

Benham, F., E. Walsh, and R. Obregon. 2015. “Evaluating Commodity Exposure Opportunities.” Meketa Investment Group.

Bessembinder, H., et al. 1995. “Mean Reversion in Equilibrium Asset Prices: Evidence from the Futures Term Structure.” *The Journal of Finance* 50 (1): 361–375.

Bessler, W., and D. Wolff. 2015. “Do Commodities Add Value in Multi-Asset Portfolios? An Out-of-Sample Analysis for Different Investment Strategies.” *Journal of Banking and Finance* 60:1–20.

Bhardwaj, G., G. Gorton, and G. Rouwenhorst. 2015. “Facts and Fantasies about Commodity Futures Ten Years Later.” NBER Working Paper 21243, June.

Bjornson, B., and C. A. Carter. 1997. "New Evidence on Agricultural Commodity Return Performance under Time Varying Risk." *American Journal of Agricultural Economics* 79 (3): 918–30.

Blitz, D., and W. De Groot. 2014. "Strategic Allocation to Commodity Factor Premiums," *Journal of Alternative Investments* 17:103–15.

Bodie, Z., and V. Rosansky. 1980. "Risk and Return in Commodity Futures." *Financial Analysts Journal* 36:27–39.

Booth, D. G., and E. F. Fama. 1992. "Diversification Returns and Asset Contributions." *Financial Analysts Journal* 48 (3): 26–32.

Burkart, D. W. 2006. "Commodities and Real-Return Strategies in the Investment Mix." CFA Institute.

Chambers, D., and J. Zdanowicz. 2014. "The Limitations of Diversification Return." *Journal of Portfolio Management* 40 (4): 65–76.

Chong, J., and J. Miffre. 2008. "Conditional Return Correlations between Commodity Futures and Traditional Assets." EDHEC Working Paper.

De Groot, W., D. Karstanje, and W. Zhou. 2014. "Exploiting Commodity Momentum along the Futures Curves." Robeco Asset Management and Erasmus University, Rotterdam.

Dow Jones Indices. 2012. "The Role of Commodities in Asset Allocation." S&P Dow Jones, LLC.

Erb, C. B., and C. R. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal* 62 (2): 69–97.

Fama, E. F., and K. R. French. 1988. "Business Cycles and the Behavior of Metals Prices." *Journal of Finance* 43 (5): 1075–93.

Froot, K. 1995. "Hedging Portfolios with Real Assets." *Journal of Portfolio Management* 21 (4): 60–77.

Gorton, G. B., and K. G. Rouwenhorst. 2006. "Facts and Fantasies about Commodities Futures." *Financial Analysts Journal* 62 (2): 47–68.

Gospodinov, N., and S. Ng. 2013. "Commodity Prices, Convenience Yields, and Inflation." *Review of Economics and Statistics* 95 (1): 206–219.

Greer, R. J. 1978. "Conservative Commodities: A Key Inflation Hedge." *Journal of Portfolio Management* 4 (4): 26–29.

———. 2000. "The Nature of Commodity Index Returns." *Journal of Alternative Investments* 3 (1): 45–52.

Idzorek, T. 2006. "Strategic Asset Allocation and Commodities." White paper, Ibbotson Associates.

Jensen, G. R., and J. M. Mercer. 2011. "Commodities as an Investment." *Research Foundation Literature Reviews* 6 (2): 1–33.

Kat, H., and R. C. Oomen. 2006. "What Every Investor Should Know about Commodities, Part I: Univariate Return Analysis." City University London, Sir John Cass Business School and Deutsche Bank AG, University of Amsterdam.

Miffre, J. 2015. "Long-Short Commodity Investing: A Review of the Literature." EDHEC Business School.

Miffre, J., and G. Rallis. 2007. "Momentum Strategies in Commodity Futures Markets." *Journal of Banking and Finance* 31 (6): 1863–86.

Mou, Y. 2011. "Limits to Arbitrage and Commodity Index Investment." PhD dissertation, Columbia University.

Pringle, A., and T. Fernandes. 2007. "Relative-Value Opportunities in Energy and Agriculture." In *Intelligent Commodity Investing*, edited by H. Till and J. Eagleeye, 313–39. London: Risk Books.

Satyanarayan, S., and P. Varangis. 1994. "An Efficient Frontier for International Portfolios and Commodity Assets." Policy Research Working Paper 1266, World Bank.

Stulz, R. 1996. "Rethinking Risk Management." *Journal of Applied Corporate Finance* 9:8–24.

Tang, K., and W. Xiong. 2012. "Index Investment and the Financialization of Commodities." *Financial Analysts Journal* 68 (5): 54–74.

———. 2008. "Case Studies from Commodity Derivatives Debacles." Working paper, EDHEC Risk and Asset Management Research Center.

Till, H., and J. Eagleeye. 2005. "Commodities: Active Strategies for Enhanced Returns." *Journal of Wealth Management*, Fall, 42–61. Also in *The Handbook of Inflation Hedging Investments*, edited by R. Greer, 127–57. New York: McGraw-Hill.

Willenbrock, S. 2011. "Diversification Return, Portfolio Rebalancing, and the Commodity Return Puzzle." *Financial Analysts Journal* 67 (4): 42–49.

CHAPTER 24

Accessing Commodity Investment Products

This chapter discusses how investors can obtain access to the risk and returns of commodity-related investments. The **return to commodity beta** may be defined as the return from direct exposure to changes in commodity prices, which result from holding a passive long position in a commodity. For example, to deliver oil-based commodity beta, an investor can use a number of investment vehicles that capture the price change in a particular oil-based commodity investment. Until the introduction of futures-based investment products, indirect investments (e.g., equity ownership of firms specializing in direct commodity market production) have been the principal means by which many investors obtain exposure to this asset class. Investment through equities of commodity-based firms, however, mixes equity beta with commodity beta. The matter is further complicated by the complex commodity exposures and degree of commodity price risk hedging by commodity-based corporations. However, within the past decade or two, the number of investable commodity indices and commodity-linked investments has increased dramatically. This chapter provides an outline of the wide range of investment products that are currently available to acquire commodity beta exposure. Since most commodity investments are in some way linked to commodity indices, we discuss the characteristics of commodity indices and enhanced commodity indices, as well as the calculation of their returns. Finally, the chapter provides an overview of some of the major commodity indices upon which the returns of many commodity investment vehicles are based.

24.1 DIRECT PHYSICAL OWNERSHIP OF COMMODITIES

Most investors avoid holding physical commodities because storage can be cumbersome and expensive, and they prefer to avoid committing large amounts of capital. For example, barrels of oil require a storage tank as well as transportation from the purchase site. In fact, some commodities are perishable (e.g., many agricultural and livestock commodities), making them virtually impossible to store for an extended period. Precious metals are an exception. Investors often hold precious metals in the form of bullion or coinage. Gold is especially easy to store, and gold investors have historically preferred to hold physical gold rather than gold derivatives. In short, real assets such as physical commodities require a degree of active management to maintain their value. Furthermore, as discussed in Chapter 22, research suggests that over the long term, the prices of physical (i.e., spot) commodities have not kept up with inflation, which is in stark contrast to the returns generated by other forms of commodity investment, such as commodity futures and indices.[1] As a result, most commodity investments are made through derivatives contracts, such as futures contracts, forward contracts, swaps, options, and swaptions. This allows the owner to benefit from price changes in the commodity without the need to store it.

24.2 INDIRECT OWNERSHIP OF COMMODITIES

The most common method of obtaining commodity exposure is through **indirect commodity investments** involving equity, fixed income, and derivative instruments. Most investors actually have this type of exposure embedded in their traditional investment portfolios. These exposures can be as simple as traditional stock or bond investments in companies that are involved in the production, transportation, and marketing of commodities, or they can be one of a number of more specialized commodity-based investments, as outlined in the following pages. A great deal of institutional investment in commodities takes the form of index-based investments. The primary vehicle used by institutional investors for exposure to commodity indices is a commodity index swap.

24.2.1 Commodity Index Swaps

A **commodity index swap** is an exchange of two cash flows in which one of the cash flows is based on the price of a specific commodity or a commodity index, whereas the other cash flow is fixed. Similar to other types of swaps, the contract specifies a term for the swap, the frequency of the payments, and leverage (both direction and magnitude). The buyer of the swap will agree to make fixed payments on specific dates in exchange for payments that are tied to the value of a specific commodity or commodity index. Similar to a futures contract that has a market value of zero when the contract is entered into, the initial market value of a commodity swap is zero as well. The size of the fixed payments is determined such that the initial market value of the swap is zero.

A commodity swap can be replicated using a portfolio of futures contracts. Therefore, it is natural to ask why investors use swaps when they can use futures contracts. The primary reason is that the swap gives investors significant flexibility. The underlying commodity index can be chosen to meet the specific needs of the investor when there may not be exchange-traded futures contracts that would meet those needs. Also, if the swap is long-term, futures contracts of long maturity may not be available and may not be liquid enough. Further, investors sometimes prefer this structure because it allows them to maintain control of their cash. While most indices include a collateral return equal to that of Treasury bills, a majority of investors can achieve higher collateral returns by managing the cash themselves. The cash can also be used in portable alpha strategies or, if desired, held in Treasury bills.

Commodity index swaps are preferred by institutional commodity investments because swaps are competitively priced, with multiple dealers making markets in swaps on several major commodity indices. Competition among dealers provides multiple counterparties to spread the default risk that stems from over-the-counter (OTC) swap transactions. The principal drawback of OTC swaps is that only relatively few investors have direct access to this market, as it is limited to institutional investors. Also, the secondary market for commodity swaps is not liquid, so early termination or modification of swap agreements typically requires negotiating with the counterparty. Swaps also experience greater counterparty risk than do the commodity futures markets. It is important to note that in recent years, certain OTC products have been moved to exchanges, and this trend is likely to continue.

24.2.2 Public Commodity-Based Equities

Owning the equity of a firm that derives a significant part of its revenue from the sale of physical commodities is a straightforward way of gaining exposure to commodities. However, not all of these investments have returns that are highly correlated with specific commodities. For example, the returns of integrated oil companies (like Exxon) and gold-mining companies (like Goldcorp) tend to be correlated with the prices of crude oil and gold, respectively, but the returns of food companies typically have only a modest correlation with the prices of grains and livestock. There is evidence that commodity producers engage in selective hedging, in which they actively alter their hedge ratios based on their views of future commodity prices.[2] Hedging activity, to the extent that it is not disclosed in public filings, can result in unpredictable commodity exposures for indirect commodity investors. Therefore, indirect commodity investments derive at least some of their returns from the active management of commodities.

Using equities for commodity exposure subjects the investor to significant stock market exposure, a risk that can result in a decoupling of realized returns from those sought by exposure to a particular commodity. In addition to the underlying commodity and stock markets risks, an investor who takes a position in the equity securities of companies engaged in the sale of physical commodities is also subject to each company's underlying business risk.[3]

In various studies, an index of commodity-producing equities had similar long-run returns as an index of commodity futures contracts, and the long-run volatility of the commodity futures index was lower than that of the commodity equity portfolio.[4] However, the returns and volatility tell only part of the story. The same studies have shown that the portfolio of commodity-based equities had a higher correlation with the S&P 500 than with the portfolio of commodity futures. Commodity company stocks behave more like equities than like commodity futures and thus, in terms of performance, are not a good proxy for an investment in commodity futures. This result is particularly relevant for investors who are interested in accessing commodity exposure to (1) diversify traditional stock and bond portfolios, or to (2) gain exposure to replicate a specific commodity's performance.

Exhibit 24.1 displays the scatter plot of monthly returns on the S&P GSCI against the monthly returns on the S&P North American Natural Resources Index, which measures performance of equities of North American companies belonging to this sector. We can see that there is indeed a positive relationship between the two indices. The slope of the regression is 0.65, which indicates that for one percentage increase in the S&P GSCI, the equities of commodity-producing firms increase by

EXHIBIT 24.1 S&P GSCI versus S&P North American Natural Resources Index, September 1996–January 2016
Source: Bloomberg and authors' calculations.

0.65% on average. The *r*-squared of the regression is about 44%, which is statistically significant but also indicates that only 44% of the return volatility on these equities can be explained by changes in the commodity futures index. If one were to break down this chart for different sub-periods, it would be seen that the relationship is much stronger when commodity prices are strongly increasing. For example, the *r*-squared is about 72% and the slope is about 0.75 for the period 2003–8, when there were strong positive trends in commodity prices. On the other hand, the relationship is much weaker during 1996–2002 and 2008–15. These results indicate that an active allocation to equities of commodity-producing firms could capture most of the beta of commodity prices when such exposures are most desirable—that is, when commodity prices are rising.

24.2.3 Bonds Issued by Commodity Firms

Similarly, investors can gain exposure to commodities through owning bonds issued by firms that derive a significant portion of their revenue from the sale of commodities. However, if a company is rated as investment grade, the yield on these bonds generally has a lower sensitivity to the commodity market in which the company is involved. Specifically, the bond price will fluctuate less with the value of the commodities that provide the basis of the firm's revenue than with economy-wide changes in interest rates and general credit spreads. This is a result of the company's bonds having very low default risk; therefore, the bond's performance is not as sensitive to commodity prices, meaning there is little chance that any related detrimental impact to cash flows from falling commodity prices would undermine the company's ability to service outstanding debt payments. In contrast, for high-yield bonds, where the probability of default is relatively elevated, the price of the bonds will tend to be much more highly correlated with the commodity price, which drives the firm's revenues and associated operating cash flows.

24.2.4 Commodity-Based Mutual Funds and Exchange-Traded Products

Other ways to gain exposure to commodities are through commodity-based mutual funds or exchange-traded products. These vehicles typically use one of four methods to acquire commodity exposure:

1. Commodity index funds or swaps
2. Equities of commodity-based companies
3. Physical commodities
4. Commodity futures

Commodity-based mutual funds are available in different active and passive structures, including commodity-based equity funds and passive commodity index tracking funds, which benchmark to commodity price indices as well as stocks of commodity producers. While index mutual funds, which track commodity indices, are generally considered passive funds, they have opportunities to attempt to outperform their benchmarks by actively managing the collateral portion of the investments. Commodity-based mutual funds generally have fees and cost structures similar to other types of mutual funds.[5]

Commodity-based exchange-traded funds (ETFs) generally offer exposures similar to those of commodity mutual funds. Long-only commodity index-based ETFs may track broad-based commodity indices (such as the iShares S&P GSCI Commodity-Indexed Trust) or commodity sub-indices (such as the PowerShares DB Energy Fund). These ETFs generate their returns directly—by holding commodity futures contracts—or indirectly (such as through swaps). ETFs are also available that track the spot price of individual commodities, such as the SPDR Gold Shares (GLD), which tracks the price of physical gold. Rather than tracking commodity futures, the GLD ETF holds gold bullion (which is more a function of gold investment; many other specific commodity ETFs hold their exposures in the form of derivative contracts). While most commodity ETFs track commodity prices (based on commodity futures or spot commodities), some ETFs track the equity prices of firms engaged in commodity production. For example, the Market Vectors Agribusiness ETF tracks the stocks of agribusiness firms.[6] It is worth noting that equity-based ETFs may provide a less pure exposure to commodity prices than derivatives-based commodity ETFs, since their performance is influenced by factors that drive the equity markets as well as factors that drive commodity prices.[7] Some commodity-related ETFs offer leveraged or leveraged inverse commodity exposures.

The most obvious difference between commodity mutual funds and commodity ETFs is that ETFs trade on organized exchanges, like traditional equities, and generally charge lower fees than mutual funds do, while occasionally offering substantial short-term levels of leveraged and/or inversed exposure to the underlying commodity's performance.

ETFs, including commodity-linked ones, may trade at a premium or a discount relative to the net asset value of the fund. Large premiums or discounts are rare because they can be arbitraged away. If the ETF is trading at a premium, authorized participants will purchase the securities in the ETF basket, exchange them for ETF units, and then sell the units on the exchange to earn a profit. If the ETF is trading at a discount, authorized participants will buy the ETF, exchange the ETF for units of the securities that are in the ETF basket, and sell the securities in the market to earn a profit.

Since premiums and discounts are arbitraged away, the tracking error between the ETF's performance and the performance of the benchmark that it is supposed to replicate is typically very small and can be explained by fees. However, because most commodity ETFs do not invest in physical commodities or securities that directly track spot prices, there could be large tracking errors between the actual performance of the ETF and the benchmark that it is supposed to track. As these ETFs use futures and swaps to replicate the benchmarks, they may have larger and more varying tracking errors compared to equity or fixed-income ETFs. Generally, commodity ETFs that use full replication, such as the iShares Silver Trust (SLV) and GLD, have the lowest tracking errors, followed by ETFs that use futures, followed by swap-based ETFs.[8]

24.2.5 Public and Private Commodity Partnerships

Some commodity investment vehicles have structures similar to private real estate partnerships or real estate investment trusts (REITs). Private commodity partnerships offer long-term ownership of real assets used in mining and energy markets. Whereas equity real estate vehicles own and operate income-producing properties, commodity partnerships might own the extraction rights to a natural gas field, a pipeline, railcars, storage facilities, or refining operations related to natural resource extraction. Partnerships are organized as pass-through entities. The income from ownership of the assets is distributed to the partners, typically without taxation at the corporate level. Income can be in the form of rental income—for example, if the partnership owns a pipeline—or from the sale of such commodities as crude oil or natural gas. In either case, the income paid by the partnership is usually correlated with the price of some commodity, so investors see partnerships as a highly correlated substitute for direct commodity investments.

In the United States, a master limited partnership (MLP) structure is a tax-efficient structure that can be used to gain access to sources of returns that are correlated to certain parts of the commodity sector. An MLP is a partnership that has its shares (called "units") traded on a public exchange, such as the NYSE or NASDAQ. As partnerships, MLPs are pass-through entities for tax purposes, meaning they do not pay taxes at the corporate level. As currently defined by the U.S. Tax Code, MLPs are required to generate at least 90% of their income from activities with "qualified sources," such as depletable natural resources.[9]

Such activities include oil and gas exploration and production, mining, gathering and processing, refining, compression, transportation, storage, marketing, and distribution. MLPs operate in a number of natural resources–related businesses and have been popular vehicles for investment due to their tax-advantaged high-distribution payout structure and, in certain cases, cash flows backed by long-term contracts. While REITs have statutorily required distribution minimums, MLPs do not. MLPs have the ability to grow and thus increase distributions through additions to their asset bases by acquisition or development.

An MLP's ownership consists of a general partner (GP) and limited partners (LPs). LPs provide capital but have no role in managing or operating the MLP and have limited voting rights. They are, however, entitled to receive cash distributions, and their units are publicly traded, allowing for liquidity. In contrast, the GP tends to hold a small stake (e.g., 2%) but has full management responsibility and control of the business. Many MLPs operate what may be described as a "toll-road" business model, meaning they receive a fee for handling the customers' product without taking ownership of the commodity. MLPs can have long-term contracts with their customers, often with attractive features such as "take-or-pay" and inflation escalators that help provide cash flow stability and limited commodity price exposure. MLPs typically operate in asset intensive businesses with high barriers to entry, which can also help ensure their cash flow stability.[10]

As with any investment, there are unique risks related to investing in MLPs. Some of these risks are due to the legal structure of MLPs, and some are borne of the market in which they operate. MLPs, particularly in times of market stress, can demonstrate equity-like volatility and drawdowns. The majority of MLPs continue to be held by retail investors who may react to negative news by selling their positions. Historically, MLPs have increased their distributions in aggregate. However, individual MLPs can change and, in some cases, decrease their distributions for strategic, competitive, or other business reasons. The MLP marketplace remains small compared to domestic equities and bonds. Investors with larger portfolios may experience difficulty in efficiently building or reducing their positions due to limited trading volumes. Much of the market's investable value is represented by a limited number of MLPs.

A key attraction of MLPs has been their comparatively high yields. An overall rise in interest rates could diminish MLPs' appeal if they are not able to generate a commensurate growth in distributions. MLPs typically distribute a very high percentage of their free cash flow and, as such, need to regularly access the capital markets for debt and equity to finance their growth. Equity capital raises could be dilutive to existing unit holders.

24.2.6 Commodity-Linked Investments

A commodity index-linked note is a debt instrument that pays a return linked to the performance of a commodity (e.g., natural gas or aluminum) or a basket of commodities over a defined period. The note may not pay any interim coupons, with the only payoff distributed on the maturity date, when the note pays the initial principal amount plus return, if any, based on the percentage change in the underlying commodity. The market for this product is somewhat more expensive than that for swaps but offers a number of advantages, including much smaller minimum investments. Also, index-linked notes may include other features, such as principal protection, that could appeal to small investors. Index-linked notes appeal primarily to investors who prefer to hold bonds, often for regulatory purposes. Many investors are obligated to own securities and have difficulty owning futures contracts. Furthermore, futures contracts and swaps require frequent margin and/or collateral postings. Index-linked note structures do not require collateral, since the structure is already fully collateralized. Finally, there is a secondary market for index-linked notes. A swap contract cannot generally be transferred, but ownership of notes is easily transferred.

A **commodity exchange-traded note** (ETN) is a debt instrument that is traded on an exchange but is different from an ETF; instead of holding an independent basket of assets, the ETN is a note issued by a financial institution that promises to pay ETN holders the return on some index over a certain period of time, and then return the principal of the investment at maturity. This means that if the company faces financial distress and it is not able to meet its obligations, ETN holders could lose either a portion or all of their investments. The credit risk exposure associated with ETNs is generally higher than the counterparty risk inherent in a commodity index-linked swap. In the case of the latter, swap counterparties post collateral at regular intervals, so counterparty risk is usually no more than a few percentage points of the total notional size of the swap.

The primary advantage of an ETN is that the issuer promises to pay exactly the return on some index, minus fees and expenses; therefore, there is little tracking error. Further, some ETNs are able to deliver the returns of a particular index that may not be available in an ETF framework. For example, the underlying securities could consist of derivative products, which in some instances may not be held in an ETF structure. Finally, in some jurisdictions, investors may face different tax treatment depending on whether they choose to invest in ETFs or ETNs.[11] ETNs are frequently referred to as prepaid forward contracts. **Prepaid forward contracts** are fully collateralized forward contracts for delivery. Thus, a commodity exchange-traded note is economically equivalent to a prepaid forward contract on the index value. Commodity indices have a high turnover rate, so profits and losses are usually treated as short-term capital gains and are therefore taxed as ordinary income, at a relatively high rate. However, since an ETN is based on the index value (rather than futures contracts), an ETN may qualify for long-term capital gains tax treatment (at a lower rate) if it is held for a minimum period of time, which is often defined as more than one year.

Long-maturity index-linked futures contracts on several commodity indices are also available. These are economically equivalent to ETNs, except that they are listed on futures exchanges rather than stock exchanges. The key difference between these contracts and typical commodity futures contracts is that long-maturity index-linked futures contracts are fully margined, meaning the trader must post 100% collateral at the time of investment (although market makers and other commercial users of these contracts are allowed to post much smaller margin amounts). Long-maturity index-linked futures contracts offer one of the least expensive and least leveraged solutions for retail investors. Furthermore, there is less counterparty credit risk associated with futures exchange trades than with ETNs.

24.2.7 Commodity-Based Hedge Funds

Hedge funds are active players in the commodity markets. Before 2000, commodity hedge funds were primarily focused on absolute return and relative value strategies. However, in response to the rise in institutional investment in commodities, a number of firms have launched long-biased commodity hedge funds. There are two categories of long-biased funds.

The first group is similar to long-only equity funds, in which managers typically hold long positions in futures contracts rather than physical commodities. Funds in this group look to identify undervalued commodities to purchase and hold in unhedged portfolios. These funds have been successful in attracting assets that would otherwise be placed in commodity index funds. Active managers may be able to avoid unfavorable rolls and to overweight or underweight commodities based on fundamental, quantitative/statistical, or technical models. The fund managers are typically benchmarked to a particular index, with compensation schemes tied to their performance relative to the fund's benchmark. Lockups are short, and both liquidity of the investments the fund makes and transparency are higher compared to other hedge funds.

The second group of long-biased funds is involved in the physical markets. These funds are engaged in the purchase, storage, and transportation of commodities. Because they also limit their use of hedging, their returns are influenced by the direction of commodity prices. These managers tend to specialize in particular commodities or commodity sectors. Investors in these funds are seeking returns generated through a combination of active management (alpha) and commodity beta. Similar to other hedge fund styles, investments with these managers are fairly illiquid, often requiring long notice periods prior to redemptions. Competitors include trading companies, shipping firms, and the commodity trading desks of major financial firms.

Hedge funds that focus exclusively on commodity investments are by no means restricted to long-only funds. Hedge funds are familiar with commodity investments and attempt to provide pure alpha, pure beta, and alpha-plus-beta options in various hedge fund products. Hedge fund alpha is available through a number of methods, including cash management, long/short positions, and instrument choice.[12] In a hedge fund, there is no guarantee that a manager will maintain a long position in commodities, as he may choose to take short positions or to hedge some or most of the fund's commodity exposure depending on his view.[13] Commodity hedge funds also use market-timing strategies.

Historically, the correlation between commodity hedge fund returns and the returns of the underlying commodities has been low. Exhibit 24.2 displays the relationship between HFR's index of hedge funds that specialize in energy and basic commodities and the S&P GSCI. We can see that there is a positive, but relatively weak, relationship between the two indices, a point that is underscored by both the

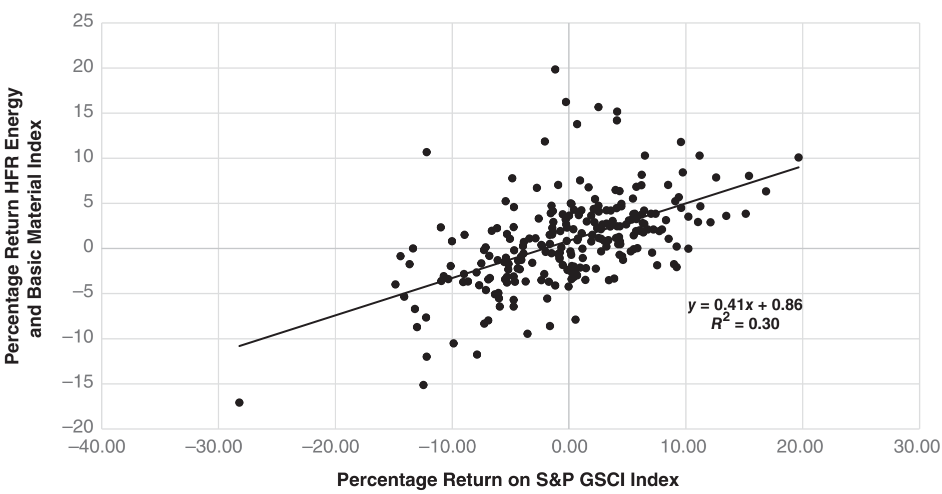


EXHIBIT 24.2 HFR Energy and Basic Materials Index and S&P GSCI, January 1990–January 2016
Source: Bloomberg, HFR, and authors' calculations.

regression's weak slope and low r-square. It is important to note that while there is a positive relationship between the commodity hedge fund index and the commodity index, the same positive relationship does not exist between average commodity hedge funds and the commodity index. In fact, a study of individual hedge funds shows that since 2006, the average correlation between commodity hedge funds and the S&P GSCI is not significantly different from zero.

24.2.8 Commodity Trade Financing and Production Financing

Financing for the production, storage, and transportation of commodities has historically been provided by merchant banks, vertically integrated commodity firms, and investment banks. A number of private investment funds have also been organized to compete in the high-margin segments of this business.

The strategies related to financing physical commodities can be executed in a number of ways. Most commonly, investment funds will provide financing for the extraction or shipping of bulk commodities, with the commodities pledged as collateral for the loans. Other strategies involve purchasing commodities for future delivery directly from the producers. In effect, the producers are borrowing money that will be repaid with commodities. This provides working capital to the producers and an effective hedge for producers against a decline in the price of the commodity. For investment funds, these transactions typically have higher margins than are available simply by purchasing commodity futures, as they are custom transactions that frequently involve operational, credit, and political risk. Many commodity producers operate primarily in parts of the world with elevated political risks. Purchasing commodities for future delivery from producers operating in precarious environments carries considerable risks (e.g., expropriation). These risks, and the resulting profits, have traditionally gone to multinational banks and other financial intermediaries. As recent regulations have forced financial institutions to dramatically reduce risk-weighted assets, profit opportunities—and their related risks—have grown for investment funds as they have stepped up to fill the gap.

24.3 LEVERAGED AND OPTION-BASED STRUCTURES

While most externally managed commodity investments are delivered through standard index-based structures such as swaps, ETFs, and structured notes, an array of complex structures, such as principal-protected notes, levered notes, options, and swaptions, are available to investors.

24.3.1 Leveraged and Short Commodity Index-Based Products

Most commodity index-based products have traditionally been designed to provide an unleveraged long exposure to a commodity index; however, with increasing regularity, there are a variety of ETFs and ETNs being designed to provide unleveraged short, leveraged long, or leveraged short (i.e., leveraged inverse) exposure to a variety of investable indices, including commodities. Managers of leveraged exchange-traded products primarily use swap and futures contracts to provide holders with either 2×

or 3× long or short exposure to the daily returns of an underlying index, in which the majority of products reset their exposure on a daily basis. This daily reset may result in a counterintuitive performance profile over a longer period of time. In other words, it is possible for the long-term return on a 3× long commodity ETF to be negative even when the price of the underlying commodity has increased over the same period. To see this, note that the daily return on a leveraged ETF can be expressed as follows (transaction costs are ignored):

$$R_t^L = A \times R_t + (1 - A)r \tag{24.1}$$

Here, R_t^L is the daily return on levered ETF, $A > 0$ is the degree of leverage (e.g., 2 or 3), R_t is the daily return on the underlying asset of the ETF, and r is the daily rate of borrowing (we assume that the borrowing and lending rates are the same). The value of \$1 invested in this ETF after T days will be

$$V_T = 1 \times \left(1 + R_1^L\right) \times \left(1 + R_2^L\right) \times \ldots \times \left(1 + R_T^L\right) \tag{24.2}$$

If the value of \$1 invested in the underlying commodity after T days is given by S_T, then the value of V_T can be approximated by

$$V_T \approx (S_T)^A \times \exp\left(\frac{A - A^2}{2} T \times \sigma_S^2 + (1 - A) \times T \times r\right) \tag{24.3}$$

Here, σ_S^2 is the variance of the daily return on the underlying commodity. It can be seen that for levered products with volatile underlying assets, the terms inside the exponential term will be negative, and therefore the value of the exponential term will be less than one. This means that for any specific price path of the underlying commodity, the price path of the levered EFT will be such that $V_T < (S_T)^A$. Further, if the terms inside the exponential term are negative enough, the exponential term will be significantly less than one, resulting in a long-term return that is negative (i.e., $V_T < 1$), even though the long-term return on the underlying commodity has been positive (i.e., $S_T > 1$).[14] It is important to note that for large positive values of S_T, the value of V_T will be extremely large, offsetting the values of those sample paths where values of V_T are smaller than one. The end result is that holding the degree of leverage constant, the average value of V_T is unaffected by changes in the volatility of the commodity.

24.3.2 Leveraged Notes

Index-linked notes may offer leveraged exposure to commodity indices as well. These are referred to as leveraged notes. A **leveraged note** is an indexed note that offers leveraged exposure (e.g., 3×) to a specified commodity index. Because these notes can lose 100% of their value when a daily decline of more than 33% in the index is experienced, the issuer and investor can be viewed as having or needing options. The issuer typically purchases option protection against further declines in the commodity. The investor benefits from implicit limited liability of the note, since any loss on the note cannot exceed the amount invested. The main benefit, which also serves as the primary risk, of leveraged notes is the high daily commodity exposure. In a

3× leveraged note, each $1,000 invested controls $3,000 of commodity exposure, so speculators can add significant leverage to their exposure to daily returns on a targeted commodity index.

24.3.3 Principal-Guaranteed Notes

Many investors prefer a form of index-linked notes referred to as principal-guaranteed notes. **Principal-guaranteed notes** are structured products that offer investors the upside opportunity to profit if commodity prices rise, combined with a downside guarantee that some, potentially all (depending on the note's terms), of the principal amount will be returned at the maturity of the structure. Notes containing so-called contingent protection clauses will not return the investor's principal unless the stated contingencies are met. Also, in the event the issuer, or counterparty, of a particular note files for bankruptcy, the note's purchasers essentially become unsecured creditors and risk losing up to all of their investment.[15]

These structures are popular with retail investors but have generated little interest from institutional investors. There are two common structures for principal-guaranteed notes. The first structure, and the most common, is the **cash-and-call strategy** or **participation note**, in which the principal guarantee comes from the issuer purchasing maturity- and principal-matched zero-coupon bonds, while the commodity-linked upside exposure comes through the issuer purchasing call options. For instance, assume that an investor wants exposure to commodities but wants to avoid a loss of principal. The capital committed by the investor is $1,000. The issuer can create a principal-protected product with exposure to commodities by following these steps. First, to protect the principal, a five-year, $1,000 zero-coupon bond yielding 5% is purchased for $783.53, which leaves $216.47 for the purchase of call options on the benchmarked commodity index. Next, upon the maturity of the bond, the $1,000 principal is returned to the investor. If the note is invested in Treasuries, then there is essentially no risk of default on the part of the issuer. If the zero-coupon bond is a liability of the issuer, then there is a chance that the principal will not be fully realized. In this case, the risk of the principal and the performance of the product will reflect the credit quality of the issuer. Finally, if the commodity price has increased sufficiently and the call option expires in-the-money, the investor will receive all or only a portion of the gains. Otherwise, the investor will receive the principal.

The second, less common, structure for principal-guaranteed notes is a dynamic strategy, or constant proportion portfolio insurance (CPPI), which varies the size of the commodity investment based on the cost of insuring the principal guarantee. Using the previous example, the initial price of the zero-coupon bond is $783.53, which is set as the floor value. This floor value changes over time, as rising yields reduce the cost of the zero-coupon bond, while the passage of time increases the cost of the zero-coupon bond, as there is less time remaining to earn interest. In the dynamic strategy, commodity investments are held as long as their value exceeds the cost of purchasing the zero-coupon bond to insure the portfolio. Some strategies will have greater investments, such as 100% of face value invested in commodities when there is a greater distance between the portfolio value and the floor, and smaller investments, such as 30% commodities, when the value of the portfolio is moving closer to the cost of the zero-coupon bond. Once the commodity portfolio has posted sufficient losses to reach the value of the zero-coupon bond floor, all commodity assets are liquidated and the principal-guaranteed note will hold 100% of the remaining assets in the zero-coupon bond until maturity. After the floor value has been touched and the portfolio has been completely moved into zeros, the note will no longer have any exposure to the underlying commodity index.

24.4 COMMODITY INDEX BASICS

Commodity indices are an effective and efficient means for gaining access to benefits from the investment performance of commodities. A commodity index is a group of commodity futures contracts that replicates the performance of a basket of closely related commodities (e.g., precious metals, agriculture and livestock, and energy) or even all traded commodities if the index is broad based. Commodity indices provide returns comparable to passive long positions in listed futures contracts. In order to gain long-term exposure, commodity futures contracts must be rolled from one maturity to the next before the expiration of the front-month contract. They attempt to replicate the returns one would generate through holding long positions in agricultural, metal, energy, and/or livestock futures contracts without requiring the investor to actively manage the relative exposures. The first generation of commodity indices, which includes the S&P GSCI, are long-only, are infrequently rebalanced, and ignore the term structure of forward prices in their weighting method. Second- and third-generation enhanced commodity indices were developed in an attempt to improve returns by providing unique roll or weighting methods, rebalancing more frequently, or implementing rules-based trading strategies that use signals derived from futures market characteristics, such as those related to momentum, mean reversion, shape of the term structure, and time remaining to expiration.

The number of commodity indices available to investors has grown noticeably over the past 25 years. A wide variety of commodity indices and sub-indices, both traditional (i.e., first-generation) and enhanced (i.e., second- and third-generation), form the basis for several publicly tradable instruments. While all indices offer a diversified exposure to commodity markets through the use of commodity futures contracts, they have differing methodologies, which include commodity selection criteria, rolling procedures, rebalancing strategies, and weighting schemes. A commodity index's performance is a function of these methodology choices and the rules-based signals previously mentioned (if second- or third-generation only), all combined with the ever-changing market prices of the index's constituents. As a result, commodity index performance can vary across indices and market environments. To reiterate, the methodology of the index will have an impact on volatility and return levels.[16]

24.5 EIGHT SOURCES OF COMMODITY INDEX RETURNS

Commodity indices can benefit from multiple sources of returns, many of which tend to be uncorrelated. As explained in the following pages, this may include one or more of the following eight sources:

- Commodity beta return
- Roll return (and roll schedule)

- Spot return
- Dynamic asset allocation
- Diversification
- Commodity component weights
- Rebalancing due to maturity
- Collateral return

Factors that impact the performance of the spot price, such as momentum and mean reversion, will also impact index performance.

- Commodity beta: For a given market, commodity beta can be defined as the return to holding a specific commodity's current front-month futures contract until the contract roll date, and then rolling the exposure to the next active front-month futures contract. With respect to liquidity and transparency, this is often the most efficient way to hold commodities and thus is the benchmark against which other commodity investment strategies are measured.
- Roll return: Roll return, or roll yield, is the profit or loss from holding a futures contract due to the change in the basis (where basis equals spot price minus futures price). It can be helpful to think of roll return as that which is obtained from rolling up or down a commodity's forward curve from the future toward the present with the passage of time. Roll returns tend to be positive when the commodity forward curve persistently slopes down (backwardation), and tend to be negative when the forward curve persistently slopes up (contango). This is a practical result of a trader holding a position in a particular futures contract, which tends to converge, or roll, toward the spot price as the maturity date of the contract approaches and the impact of the cost-of-carry and convenience yield pricing components diminish. Some commodities, such as gold, nearly always have a negative roll yield. The cost of carry for gold has been larger than the convenience yield, so the forward curve has been upward sloping. Difficult-to-store commodities and those that are subject to supply or demand shocks tend to have larger convenience yield and are more likely to be in backwardation and therefore generate positive roll returns.[17] The weighting scheme of an index can have a significant impact on its roll return. It is worth noting that there are a number of reasons to be wary of designing an index strictly on the basis of roll return:
  1. Roll return for a particular commodity changes over time with changes in interest rates, storage costs, and convenience yields. A commodity's forward curve may flip from backwardation to contango with changes in cost of carry.
  2. Commodities in backwardation typically exhibit high volatility due to their low inventory levels.
  3. Focusing strictly on roll return may negatively impact the degree of diversification of the commodity index, as commodities that tend to all be in backwardation or all in contango may share significant correlation with one another.
- Spot return: The difference between the excess return of an index and the roll return is referred to as the spot return. In other words, after adjusting for the component of return attributable to changes in the futures basis, the remaining component of return is logically due to changes in the spot price over the relevant period. Since commodities tend to mean-revert over longer periods of time, spot

return is expected to contribute little to total returns over longer periods. Of course, there can be periods of a commodity supercycle, during which spot (and futures) prices rise significantly.[18]

- Dynamic asset allocation: Some commodity indices incorporate dynamic asset allocation models. A **dynamic asset allocation model** in the context of commodities is a portfolio weighting method that determines those commodities to be overweighted and/or underweighted, with adjustments in weights being made through time. Momentum is a commonly used factor, or rule, within dynamic asset allocation models. Momentum-based models overweight commodities that have recently been increasing in price, and underweight commodities that have recently been declining in price, with the comparison often in terms of their current price relative to their historical moving average over a specified time period. Another common strategy for dynamic asset allocation is mean reversion. In contrast to momentum, this strategy reduces the allocation to commodities that have increased in value, and increases the allocation to commodities that have declined in price. Typically, momentum-based asset allocation rules are based on short-term models, while mean-reversion strategies are based on price changes over a time horizon greater than one year. Term structure signals can also be used for dynamic allocation. For example, commodities that are backwardated can be overweighted to capture an anticipated roll yield, while those that are in contango can be underweighted.
- Diversification: The greater the number of commodities, the more diverse the index, making it less sensitive to price changes in a particular commodity or, if the index is broad based, commodity subsector. Diversification can be particularly relevant for broad-based commodity indices (unless specific exposure is desired), since commodity subsectors tend to have low correlation with one another, and individual commodities tend to have high volatilities in isolation. As discussed in Chapters 22 and 23, in the presence of mean reversion in commodity prices, rebalancing can result in significant improvement in the performance of an index.[19]
- Commodity weights: Higher index returns may result from overweighting commodity futures that have recently been increasing in value and underweighting those that have been falling in value (due to the momentum return factor), or the reverse may be true if mean reversion holds. However, as with any group of investments, the risk, as measured by volatility, can increase in a commodity index that allocates a large weight to a small subset of constituents. The commodity weighting scheme should incorporate a reasonable economic rationale. For example, the weights of each commodity in the S&P GSCI are based on their average global production levels over the most recent five-year period.
- Maturity: Using longer-maturity futures contracts will affect the roll return, with the impact depending on the shape of the forward curve. If the roll yield is positive, then moving to longer dated futures contracts tends to increase the roll return, but its size will depend on the slope and overall curvature of the forward curve. The impact of maturity on an index's spot return tends to be significant. Longer maturity futures prices are less sensitive to changes in spot prices of commodities, which means that longer maturity futures contracts are less volatile (and therefore an index's return should be less volatile). In addition, indices that are based on longer maturity futures contracts will have a lower

commodity beta and therefore will underperform (outperform) other indices using contracts with shorter times to expiration when there are significant increases (decreases) in spot prices of commodities. Finally, another drawback related to the use of longer-term futures contracts is that liquidity is much less in futures contracts with longer maturities, and therefore these indices may not have the capacity required by large institutional investors.

- Collateral: In total return commodity indices, the notional amount of the investment is assumed to be held in interest-bearing instruments, often short-term sovereign debt issues. For investors tracking such an index, this collateral—which can be used as futures margin—earns interest, is referred to as collateral return or collateral yield, and enhances the overall return of commodity futures-based investments. It should be noted that while T-bills have traditionally served as the base case for analyzing collateral returns, investments in commodities may be overlaid on other portions of a fixed-income portfolio, which may provide large investors with enhanced returns earned from capturing the yield spread.

24.6 ISSUES IN COMMODITY INDEX DESIGN

There are a number of decisions that need to be made when designing a commodity index, which include weighting and roll methods, as well as collateral considerations. Enhanced commodity indices (discussed in section 24.7) may modify the roll method and the choice of contracts traded in the pursuit of higher returns.

24.6.1 Value Based versus Quantity Based

Commodity indices can be value based or quantity based. A **value-based index** has fixed-component weights expressed as percentages of the value of the index. The number of futures contracts in the index changes dynamically to maintain constant value weights. A **quantity-based index** holds a fixed quantity of contracts for each commodity, so that the index weights change each day in terms of percentage of value as futures prices change. For example, the S&P 500 Index is quantity based, since the number of shares of each company in the index changes only when the index constituents are changed, whereas the percentage of the index attributable to a particular constituent changes when the market price of that constituent changes. Conversely, an index that consists of 60% stocks and 40% bonds is value based.

24.6.2 Total Return versus Excess Return

There are two types of return indices available to commodity investors. A **total return index** (discussed further and calculated in section 24.8) is a fully collateralized investment strategy, with the collateralization generally taking the form of Treasury bills. In a total return index, the overall calculation of the index return includes the cash return from the collateral (i.e., collateral yield). Generally, total return indices have returns and volatilities comparable to broad equity indices. An **excess return index** (also discussed further and calculated in section 24.8) provides returns over cash and is linked to the price movements of a basket of commodity futures contracts.[20]

24.6.3 Roll Method

The roll method can have a significant impact on the returns of a commodity index. In its simplest form, roll methodology involves two primary choices:

1. Futures curve positioning
2. Roll procedure

Futures curve positioning determines the time to expiration of a futures contract at the initiation of the position, and the length of time the contract will be held before rolling to a further-out contract. First-generation indices, such as the S&P GSCI (which has an average maturity of less than two months), generally position themselves at the near end of the futures curve and hold their contracts for a short period of time. Many of the newer indices position themselves further out on the forward curve and/or hold their positions for a longer time before rolling. Since forward curves of commodities are generally nonlinear, the impact of contango or backwardation on returns can vary significantly due to the choice of curve positioning strategy.

When the roll return is negative, passive long commodity indices will perform poorly. A number of products have attempted to address this negative roll return problem. They include better execution (e.g., changing roll dates), buying deferred contracts (to achieve better roll returns), or optimizing exposure along the curve. While these approaches may improve the index's performance in some cases, they have the disadvantage of becoming close to an active strategy and therefore contradicting the initial goal of investing in a passive product. In response, passive products whose methods are curve-neutral have been developed. An example is the JPMorgan Commodity Curve Index (JPMCCI), which implements a simple, curve-neutral approach by holding exposure along the commodity futures curve according to the open interest of each tenor. The JPMCCI also uses open interest to determine the inclusion and relative weights of the individual commodities to arrive at a total market benchmark, whose constituents are expected to grow as futures market liquidity deepens and lengthens going forward.

The actual roll procedure can have a significant impact on returns, particularly for the more popular indices, where the sheer trading volume of funds tied to the index can move markets due to insufficient liquidity. As a result, extremely large futures positions must typically be rolled over multiple days. In addition, the roll procedure is publicly available and thus entirely predictable, which can result in significant market impact—to the detriment of index returns but to the delight of alert traders who take advantage of the opportunity for profit. Many newer indices are designed to minimize the inefficiencies of their rolls. For example, an index may roll a small portion of its positions each day, resulting in a low-impact, almost continuous roll.

24.6.4 Weighting Methodology

Perhaps the largest impact on index returns comes from the choice of weighting method. The weighting method determines the degree of diversification of the index, which is obtained from the relative concentration in particular commodities or commodity subsectors contained within the index. For example, the S&P GSCI has a very high concentration in energies (about 71% at the start of 2016), whereas the Bloomberg Commodity Index (BCOM) has a 33% cap on exposure from any subsector.[21]

There are a number of input variables an index provider can use to determine individual commodity or commodity-sector weights. The primary variables are as follows:

1. World production—for example, the S&P GSCI is a quantity-based, world production–weighted index.
2. Liquidity—for example, the BCOM uses a combination of liquidity and production measures to assign weights to individual commodities.
3. Open interest.

Index weighting schemes can vary significantly in the degree of diversification they provide. For example, the Deutsche Bank Liquid Commodity Index (DBLCI) consists of only six highly liquid commodities, whereas the Diapason Commodities Index (DCI) is one of the broadest commodity indices available, with 48 components. It is worth noting that many indices (e.g., BCOM) use maximum and/or minimum limits on individual commodity or commodity-sector allocations to limit the degree of position concentration. Weighting methods can also incorporate more active weights as well as short positions for long/short or short-biased indices.

24.6.5 First-Generation Commodity Indices

Spot-based commodity indices date back to the mid-1800s (the *Economist* commodity price index). However, spot indices can be of limited use to investors, since with few notable exceptions (e.g., gold), most commodity investment is futures based. The Commodity Research Bureau (CRB) Index, established in 1957, has the longest history of all the futures-based commodity indices, but the usefulness of its long track record has been reduced by the high number of revisions in methodology and by the fact that the index was not investable in the early years. The first generation of investable commodity indices began with the introduction of the Goldman Sachs Commodity Index (now the S&P GSCI) in November 1991 and was followed by the Dow Jones-AIG Commodity Index in 1998 (renamed the Dow Jones-UBS Commodity Index in 2009, and now the Bloomberg Commodity Index, itself the result of a rebranding on July 1, 2014). The S&P GSCI and the BCOM are by far the most dominant commodity indices and represent the first generation of investable commodity indices.

Most commodity index-based investment is tied to these two indices; however, they suffer from some limitations, which may result in suboptimal returns. Both the S&P GSCI and the BCOM invest in short-maturity futures and roll these futures to the next-further-out contract over a relatively short time frame. The sheer size and concentration of the roll of these indices results in a reduction in their returns. For example, empirical evidence has been found that significant alpha can be generated by exploiting the established roll pattern of the S&P GSCI.[22] Researchers tested two **preroll strategies**, meaning that the contracts were rolled well before settlement, on 19 commodities included in the S&P GSCI, and found that rolling futures contracts before the scheduled roll period generated significant abnormal returns over the period from 2000 to 2010, with the energy sector providing the best performance. Furthermore, first-generation indices are truly passive and make no attempt to adjust their exposures in response to the shape of the forward curves or the expected returns of the commodities they invest in.

In the years following the introduction of the S&P GSCI and the BCOM, a wide variety of index evolution has occurred. Many were simply slight evolutions of the first generation, differing by degree of diversification or the focus of their concentrated exposures. For example, while the Thomson Reuters/Jefferies CRB Index (TR/J CRB) has an average maturity and roll method similar to those of the S&P GSCI, it has the largest allocation to agriculture of any widely followed, broad-based index, as of March 2009.

24.7 PERFORMANCE ENHANCEMENTS OF NEW COMMODITY INDICES

First-generation commodity indices, discussed in section 24.6.5, tend to be heavily weighted in energy and hold long-only positions in front month contracts, rolling to the second month contracts regardless of the shape of the current term structure. This practice can result in a significant drag on returns when markets are in contango, particularly when the near-term end of the term structure tends to be the steepest. **Second-generation commodity indices** attempt to enhance returns through forward curve positioning to spread the roll period across points along the forward curve, or target different segments of the curve. **Third-generation commodity indices** add yet another enhancement to second-generation commodity indices by including active commodity selection, which may be predicated on objective rules (such as using algorithms to assign weights based on specific criteria related to momentum, inventory levels, term structure signals, and so on) or could be discretionary.

24.7.1 Analysis of Performance of Commodity Index Enhancements

Unfortunately, validating the performance improvement due to the methodology enhancements of these indices has proven to be quite difficult. Commodity futures markets tend to exhibit high volatility and low correlations across subsectors.[23] As a result, the returns of commodity indices have been driven by a few commodities that have historically shown extreme positive or negative performance. This makes ex post comparisons between index methods difficult, since a high weight in a commodity that happens to perform extremely well or poorly can have an overwhelming impact on returns.

To analyze the potential improvements of enhanced commodity indices while addressing this issue, researchers compared two first-generation indices (the S&P GSCI and the BCOM) with enhanced versions of the indices. To create the enhanced versions of the indices, they used the weighting and rolling methodology of the first-generation indices as a baseline, and adjusted the weighting or time to expiration of the futures contracts. The weighting was adjusted based on a momentum signal (i.e., overweight the winners), a term structure (overweight commodities with higher roll return), or a combined momentum and term structure signal to create reweighted enhanced indices. The time to expiration was also altered to create enhanced indices that retained the baseline weights. The researchers found that all of their hypothetical enhanced versions of the S&P GSCI and the BCOM improved the returns of the first-generation indices, providing significant positive alphas. The largest improvement came from the time-to-expiration enhancement, with the longer-maturity versions exhibiting the highest outperformance. Similarly important, the enhanced indices provided effective risk diversification and inflation hedging capabilities when compared to the underlying first-generation indices.[24]

These hypothetical results are supported by actual ex post performance of a wide variety of indices.[25] These second- and third-generation indices typically outperform the first-generation indices.

24.7.2 Second-Generation Enhanced Commodity Indices

The second-generation commodity indices attempt to enhance returns through forward curve positioning. In contrast to first-generation indices, second-generation indices tend to spread the roll period across points along the forward curve or target different segments of the curve. A variety of enhanced commodity indices have evolved, which all claim to provide return improvements over the first generation of commodity indices without increasing risk or reducing hedging capability. These roll techniques may be best classified into four categories:

1. Enhanced roll techniques attempt to reduce the impact of contangoed markets by rolling into mid- to long-term contracts and holding them close to expiration. This practice reduces the frequency of rolling while also spreading the roll across less steep portions of the term structure.
2. Constant expiration techniques spread their positions within a market across a range of expirations, averaging to their target maturity. Once again, this technique allows them to avoid concentrating on the steepest portion of the term structure.
3. Implied roll yield techniques select the contract with the most attractive roll yield within an acceptable maturity range.
4. Other techniques move the contracts further out on the forward curve or use long-only momentum, mean reversion, or liquidity signals to adjust positioning for market conditions.

An example of a second-generation index is the Merrill Lynch Commodity Index eXtra (MLCX), which differs from the S&P GSCI and the BCOM in its longer average time to expiration. The MLCX rolls contracts over a period of 15 days, from the first to the 15th business day of the rolling month, and rolls from next to second-next contract (i.e., from the second contract to the third contract) instead of the more conventional from front to next contract (i.e., from the first contract to the second contract). Rolling one month ahead of the S&P GSCI and the BCOM gives the MLCX an average time to expiration of about one month longer than the BCOM and six weeks longer than the S&P GSCI.

24.7.3 Third-Generation Enhanced Commodity Indices

Third-generation commodity indices add yet another enhancement: active commodity selection, which may be based on objective rules (such as using algorithms to assign weights based on specific criteria related to momentum, inventory levels, term structure signals, and so on) or could be discretionary. In contrast to the long-only first- and second-generation indices, third-generation indices follow dynamic long-only or long/short strategies. There are four basic types of third-generation classifications:[26]

1. Momentum-based index strategies use price momentum to establish their long and short futures positions.
2. Term structure approaches attempt to capitalize fully on roll yield by taking long positions in commodities with the steepest backwardated forward curves and shorting those commodities with the greatest degree of contango.
3. Market neutral strategies balance long and short positions to maintain a net market neutral position.
4. Rule-based approaches use a combination of fundamental analysis and technical analysis to determine commodity weights.

An example of a term structure third-generation commodity index is the CYD Long-Short TR Index, which equally weights long positions in backwardated commodities and short positions in contangoed commodities.

24.8 COMMODITY INDEX RETURN CALCULATION

Commodity index return attribution can benefit from analysis of four types of returns: spot returns, roll returns, excess returns, and total returns.

24.8.1 Four Challenges with Return Attribution to Commodity Indices

Different commodity indices can have substantially different returns over the same period and with the same investment universes. Returns on commodity indices can be substantially influenced by four major factors that are not found in more traditional indices, such as buy-and-hold equity indices. Attention, therefore, needs to be paid to return attribution, especially in the context of the following four factors:

- Weights. Returns to commodity indices can differ substantially due to considerably different weighting schemes of the underlying contracts. The effects of these weighting differences are primarily captured through differences in the spot returns underlying the index.
- Rollover entry points. Commodity index returns are influenced by the time to settlement of each futures contract that is opened when a position is rolled. The returns of contracts on various settlement dates can differ due to the particular bases of the contracts.
- Rollover exit points. Commodity index returns are influenced by the timing of the initiation of rollovers (i.e., the exit points in terms of time to settlement). An extreme example is a preroll strategy versus a rollover at settlement.
- Collateral yields. Commodity index returns are influenced by differences in returns on the collateral, caused by such factors as the different exposures of the collateral to duration risk or default risk.

The index calculation primer detailed in the next section is a standardized method developed for commodity indices to assist in return attribution.

24.8.2 Index Calculation Primer

The four measures of return that are typically published by commodity index providers are spot return, excess return, total return, and realized roll return. Each of these measures of commodity performance has an important use in evaluating the performance of different commodity investment strategies. This section focuses on analysis of these return measures, along with determining three index values: spot index, excess return index, and total return index. It concludes with a practical summary of these concepts by calculating hypothetical commodity index values and returns over the course of three trading days, one of which involves a roll of the index.

The industry convention is to compute index returns according to the following method.

24.8.2.1 Computation of Spot Return The spot return represents the percentage change in the market value of futures contracts held in the index. This return generally assumes that no roll occurs. If a roll occurs, it reflects the price difference between the old and the new futures contracts, and would also incorporate any change to the quantity of contracts held by the index due to the roll. The following expression describes the spot index at time $t-1$:

$$\text{Spot index } (t-1) = w_1 \times F_1(t-1, t_1) + \ldots + w_N \times F_N(t-1, t_1) \qquad (24.4)$$

Here, $F_i(t-1, t_1)$ is the futures price of the contract i at time $t-1$, with maturity date t_1, and w_i is the number of contract i in the index. The percentage change in the spot index represents the spot return.

At time t, the spot index is calculated as a weighted sum of futures prices. However, compared to time $t-1$, the index could roll to futures contracts with longer maturity, the number of contracts for each commodity could change, or both. Any of these changes could result in realizing a spot return even if the entire forward curve remained unchanged. In such a case, the roll return would appear. (Roll return is discussed later in the section.)

The following expressions describe the spot index at time t for the three possible states of the index:

$$\text{Spot index}_t = \begin{cases} w_1 \times F_1(t, t_1) + \ldots + w_N \times F_N(t, t_1) & \text{No roll return, no change in \# of contracts} \\ w_1 \times F_1(t, t_2) + \ldots + w_N \times F_N(t, t_2) & \text{Roll return, no change in \# of contracts} \\ w'_1 \times F_1(t, t_2) + \ldots + w'_N \times F_N(t, t_2) & \text{Roll return, changes in \# of contracts} \end{cases} \qquad (24.5)$$

Here, w'_{i1} is a new number of contract i in the index. (Note: When calculating the spot return when a roll occurred between spot index values, w'_i would appear in the numerator and w_i would appear in the denominator.)

24.8.2.2 Computation of Excess Return Excess return represents the percentage change in the market value of the futures contracts held in the index at the end of the trading session but before accounting for any index changes (i.e., rolls or changes in the number of contracts). There will be a difference between spot and excess return when the index's composition changes, either because a contract is rolled forward or because the weights are changed.

$$\text{Excess Return Index}_t = \text{Excess Return Index}_{t-1} \times \left(\frac{w_1 F_1(t, t_1) + \ldots + w_N F_N(t, t_1)}{w_1 F_1(t-1, t_1) + \ldots + w_N F_N(t-1, t_1)} \right) \quad (24.6)$$

Note that the numerator and denominator of the terms appearing in the parentheses of equation (24.6) refer to spot indices at time t and $t - 1$ *before* any change in the index composition. The percentage change in the excess return index creates the excess return, which represents the return that would be generated by an investment in the commodity index before changes and, to the extent the position is collateralized, net of the collateral return.

24.8.2.3 Computation of Total Return **Total return** represents the return to a fully collateralized position in the commodity index, including the collateral return. It is calculated as the excess return plus the risk-free return, r (usually Treasury bills):

$$\text{Total Return Index}_t = \text{Total Return Index}_{t-1} \times \left(\frac{\text{Excess Return Index}_t}{\text{Excess Return Index}_{t-1}} \right) \times (1 + r) \quad (24.7)$$

The percentage change in the total return index creates the total return.

24.8.2.4 Computation of Realized Roll Return **Realized roll return** represents the recognition of the accrued gain/loss associated with rolling into higher-priced/lower-priced futures contracts and/or decreasing/increasing the quantity of contracts held. The actual return represented by roll return is accrued on a daily basis (not simply when the contract is rolled) and theoretically represents the accumulated cost of carry (net of convenience yield) over the period the contract is held, adjusted by the expected cost of carry for the newly rolled-into contract. It is worth noting that roll return as measured in index return calculations is an artificial construct, as is the corresponding component of spot returns when rolls occur (i.e., realizing a spot return when rolling even if there is no change in the forward curve). Realized roll return is typically calculated as excess return minus spot return:

$$\text{Realized Roll Return}_t = \left(\frac{\text{Excess Return Index}_t}{\text{Excess Return Index}_{t-1}} \right) - \left(\frac{\text{Spot Index}_t}{\text{Spot Index}_{t-1}} \right) \quad (24.8)$$

For most indices, rolls take place on only a few days per month, and weights change once per year; thus, on most days, the realized roll return is zero.

24.8.3 Index Calculation Example

The following example illustrates the method and calculations for a simple commodity index composed of two commodities over a three-day period (see Exhibit 24.3). The figures are hypothetical. The details of these calculations are presented in Exhibit 24.4

The second column (Day 2) in Exhibit 24.3 shows return calculations on a date when the index did not change any positions (i.e., did not roll or rebalance). In most indices, the number of contracts held for each commodity in the index is fixed, except when weights are rebalanced, typically once per year. Since there is no realized roll return in this portion of the example, the spot index uses futures with the same expiration months used in calculating the excess return and total return indices. Since there were no changes in the composition of the index, spot return and excess return are the same and the roll return is zero. Note that the weight of crude oil in the index rose. This is because the index holds a fixed position in each commodity, so with the price of crude oil rising by a higher percentage than copper fell, the portion of the index represented by crude oil increased as well.

The next column in the exhibit (Day 3) shows the same set of calculations but on a date when the index rolls. On a roll date, futures contracts nearing expiration

EXHIBIT 24.3 Example of Commodity Index Return Calculations

	Day 1	Day 2	Day 3
Contract Prices			
Crude Oil (October)	80	85	81
Copper (December)	320	310	300
Crude Oil (November)			80
Copper (March)			290
Number of Contracts			
Crude Oil	60	60	60
Copper	8	8	8
Index Weights			
Crude Oil	65.22%	67.28%	67.42%
Copper	34.78%	32.72%	32.58%
Total	100.00%	100.00%	100.00%
Interest Rate			
U.S. T-Bill	2.00%	2.10%	2.20%
Index Values			
Spot Index	7,360	7,580	7,120
Excess Return Index	10,000	10,299	9,864
Total Return Index	24,000	24,719	23,677
Index Returns			
Spot Return		2.99%	–6.07%
Excess Return		2.99%	–4.22%
Total Return		2.99%	–4.22%
Realized Roll		0.00%	1.85%

EXHIBIT 24.4 Guide to Calculations in Exhibit 24.3

	Day 1	Day 2	Day 3
Contract Prices			
Crude Oil (October)	O1	O2	O3
Copper (December)	C1	C2	C3
Crude Oil (November)			O4
Copper (March)			C4
Number of Contracts			
Crude Oil	60	60	60
Copper	8	8	8
Index Weights			
Crude Oil	(O1 × 60)/(Spot Index)	(O2 × 60)/(Spot Index)	(O4 × 60)/(Spot Index)
Copper	(C1 × 8)/(Spot Index)	(C2 × 8)/(Spot Index)	(C4 × 8)/(Spot Index)
Total	100.00%	100.00%	100.00%
Interest Rate			
U.S. T-Bill	T1	T2	T3
Index Values			
Spot Index	O1 × 60 + C1 × 8	O2 × 60 + C2 × 8	O4 × 60 + C4 × 8
Excess Return Index	10,000	(Excess Return Index) × (O2 × 60 + C2 × 8)/(O1 × 60 + C1 × 8)	(Excess Return Index) × (O3 × 60 + C3 × 8)/(O2 × 60 + C2 × 8)
Total Return Index	24,000	(Total Return Index) × (1 + Excess Return Index) × (1 + T1/360)	(Total Return Index) × (1 + Excess Return Index) × (1 + T2/360)
Index Returns			
Spot Return		(Spot Index Day 2)/(Spot Index Day 1) – 1	(Spot Index Day 3)/(Spot Index Day 2) – 1
Excess Return		(Excess Return Index Day 2)/(Excess Index Day 1) – 1	(Excess Return Index Day 3)/(Excess Index Day 2) – 1
Total Return		(Total Return Index Day 2)/(Total Return Index Day 1) – 1	(Total Return Index Day 3)/(Total Return Index Day 2) – 1
Realized Roll Return		Excess Return – Spot Return	Excess Return – Spot Return

are replaced in the index by new contracts. The changes take place after the futures markets close. In our example, crude oil for October delivery is replaced with a contract for November delivery. Also, copper for December delivery is replaced with a contract for delivery the following March. The changes are assumed to take place after the close of trading on Day 3.

On dates on which a roll occurs, the excess return index is calculated based on the original contracts (October crude and December copper), while the spot return uses the closing prices of the new contracts (November crude and March copper). Using values from our example, the spot return is lower than the excess return (–6.07% versus –4.22%). The difference of 1.85% represents the realized roll return. Note that on the next day, Day 4, both the spot and excess return indices would be calculated using the new futures contracts.

24.9 CONCLUSION

This chapter provided an outline of the wide variety of vehicles available for accessing commodity returns. Significant differences exist across the available options, though the most commonly used are index based. We also outlined many of the significant differences across commodity indices. Although first-generation indices remain the most commonly used, significant return, volatility, and correlation benefits may be provided by second- and third-generation indices, which address many of the limitations of traditional first-generation indices. Lastly, we presented a primer and an example of the relevant calculations for determining the various components of commodity index values and related returns.

NOTES

1. Gorton and Rouwenhorst (2006); Bhardwaj et al. (2015).
2. Stulz (1996).
3. Schneeweis (2006).
4. Gorton and Rouwenhorst (2006); Bhardwaj et al. (2015).
5. Burkart (2006).
6. For a description of agribusiness and categories tracked, please refer to MOO exchange-traded fund.
7. Please see discussion in section 24.2.2, or refer to Jensen and Mercer (2011).
8. See Guo and Leung (2014b); Murphy and Wright (2010).
9. Benham et al. (2015).
10. Benham et al. (2015).
11. Jensen and Mercer (2011).
12. Burkart (2006).
13. Till and Eagleeye (2005).
14. The formula will change for levered short ETFs. See Avellanded and Zhang (2009); Hill and Teller (2009).
15. SEC (2011).
16. Kazemi et al. (2008).
17. Till and Eagleeye (2005).
18. Till (2006).
19. Erb and Harvey (2006); Willenbrock (2011); Chambers and Zdanowicz (2015).

20. Blanch and Scheis (2006).
21. BCOM (2015), see Bloomberg Commodity Index.
22. Mou (2011).
23. Erb and Harvey (2006).
24. Rallis, Miffre, and Fuertes (2013); Miffre (2014).
25. Miffre (2014).
26. Miffre (2014).

REFERENCES

Avellaneda, M., and J. H. Lee. 2009. *Statistical Arbitrage in the U.S. Equity Market*. Courant Institute, New York University.

Benham, F., E. Walsh, and R. J. Obregon. 2015. *Evaluating Commodity Exposure Opportunities*. Available at SSRN.

Bhardwaj, G., G. Gorton, and G. Rouwenhorst. *Facts and Fantasies about Commodity Futures Ten Years Later*. No. w21243. National Bureau of Economic Research, 2015.

Blanch, F., and S. Scheis. 2006. *Selecting a Commodity Index*. Merrill Lynch Global Commodity Paper No. 4. Research report, Merrill Lynch.

Burkart, D. W. 2006. *Commodities and Real-Return Strategies in the Investment Mix*. CFA Conference Proceedings, CFA Institute.

Chambers, D. R., and Zdanowicz, J. 2015. "The Limitations of Diversification Return." *Journal of Portfolio Management*. 40 (4): 65–77.

Erb, C. B., and C. R. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal* 62 (2): 69–97.

European Commission. 2011. *Impact Assessment of the Common Agricultural Policy towards 2020*, Annex 4, SEC (2011) 1153 final/2.

Fuertes, A. M., J. Miffre, and A. Fernández-Pérez. 2014. "Commodity Strategies Based on Momentum, Term Structure and Idiosyncratic Volatility." *The Journal of Futures Markets* 35 (3): 274–97.

Garyn-Tal, S. 2013. "Explaining and Predicting ETFs Alphas: The Rs Methodology." *The Journal of Index Investing*. 4 (4): 19–32.

Gorton, G., and K.G. Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures." *Financial Analysts Journal*, CFA Institute.

Guo, K., and T. Leung. 2014a. "Commodity Leveraged ETFs: Tracking Errors, Volatility Decay and Trading Strategies." *The Journal of Risk*, October 31.

———. 2014b. *Understanding the Tracking Errors of Commodity Leveraged ETFs*. Columbia University.

Hill, J. M., and S. G. Teller. 2009. "Rebalancing Leveraged and Inverse Funds." *Eighth Annual Guide to Exchange Traded Funds and Indexing Innovations*, *Institutional Investor Journals* (Fall): 67–76.

Jensen, G. R., and J. M. Mercer. 2011. "Commodities as an Investment." *Research Foundation Literature Reviews* 6 (2): 1–33.

Kazemi, H., T. Schneeweis, and R. Spurgin. 2008. "The Benefits of Commodity Investment." Research report, Alternative Investment Analytics LLC.

Mou, Y. 2011. "Limits to Arbitrage and Commodity Index Investment: Front-Running the Goldman Roll." Working paper.

Miffre, J. 2014. "Comparing First, Second and Third Generation Commodity Indices." *Alternative Investment Analyst Review* 3 (2): 22–33.

Murphy, R., and C. Wright. 2010. *An Empirical Investigation of the Performance of Commodity-Based Leveraged ETFs*. Symmetry Partners and Central Michigan University.

Rallis, G., J. Miffre, and A. Fuertes. 2013. "Strategic and Tactical Roles of Enhanced Commodity Indices." *The Journal of Futures Markets* 33 (10): 965–992.

Schneeweis, T., and R. Spurgin. 2006. "Early Reporting Effects on Hedge Fund and CTA Returns." *The Journal of Alternative Investments* 9 (2): 30–45.

Stulz, R. M. 1996. "Rethinking Risk Management." *Journal of Applied Corporate Finance* 9 (3): 8–25.

Till, H. 2006. "Structural Sources of Return and Risk in Commodity Futures Investments." Working paper, EDHEC Risk and Asset Management Research Centre.

Till, H., and J. Eagleeye. 2005. "Commodities: Active Strategies for Enhanced Return." *The Journal of Wealth Management* 8 (2): 42–62.

Willenbrock, Scott. 2011. "Diversification Return, Portfolio Rebalancing, and the Commodity Return Puzzle." *Financial Analysts Journal* 67 (4): 42–49.

PART

5

Hedge Funds and Managed Futures

CHAPTER 25

Managed Futures

Managed futures is a subclass of alternative investment strategies. For these strategies, professional money managers (also known as commodity trading advisers [CTAs]) manage client assets by actively taking positions primarily in futures markets, forward markets, options and other liquid derivatives, and structured products. Using highly liquid mark-to-market contracts, they typically provide their clients access to a wide range of asset classes, including fixed income, currencies, equity indices, soft commodities, energy, and metals. They apply leverage either directly via margin or indirectly via the use of derivative products, such as futures and options. Another key feature of this strategy group is the ability to go long or short with relative ease. The flexibility and liquid nature of managed futures strategies coupled with the wide array of markets they trade provide risk and return patterns not easily accessible through traditional asset classes (such as long-only stock and bond portfolios) or other alternative investments (such as hedge funds, real estate, private equity, or long-only commodities).

25.1 THE STRUCTURE OF THE MANAGED FUTURES INDUSTRY

Managed futures is a relatively wide class of strategies that share two key features. First, they are "managed," meaning that professional money managers or commodity trading advisers (CTAs) make trading decisions on behalf of an account. "Futures" denotes the industry's primary focus on using futures contracts or similar instruments. These products are desirable for these strategies because of the transparency and reduced counterparty and credit risk associated with exchange-traded instruments. For example, futures contracts are transparent because they are standardized and highly specified and are traded on markets that pool the collateral of all participants. For these contracts, the clearinghouse takes the other side of the trade, and pooling funds reduces the counterparty and credit risks of bilateral transactions. The pooling of positions and the futures exchange structure allow for substantial reduction in the margin capital required for establishing positions in futures contracts. For example, according to the CME Group, the eurodollar contract, which is one of the most liquid, can have margins as low as USD 650, or 0.065% of a $1 million notional contract. By comparison, the required margin for a long position in an exchange-traded stock is 50%.

Investors can access the managed futures industry either by investing in a futures trading fund (via a managed account or a commingled fund) or through a commodity pool—a commingled investment vehicle that resembles a fund of funds and is managed by a **commodity pool operator (CPO)**, who invests in a number of underlying CTAs. Investments from a number of investors are pooled together and then invested in futures contracts either directly by the CPO or through one or more commodity trading advisers. CPOs may be either public or private. In the United States, the requirements for investing in public futures funds generally differ from state to state; and globally, the requirements vary from country to country. Chapter 26 will detail the different approaches for structuring investment in managed futures.

Globally, the futures trading industry has a relatively short history of regulation. In the United States, the **Commodity Futures Trading Commission (CFTC)** was initiated in 1974 as a federal regulatory agency for all futures and derivatives trading. This regulatory body was later supplemented with U.S. futures exchanges and the **National Futures Association (NFA)**, an independent, industry-supported, self-regulatory body created in 1982. In Europe and Asia, managed futures funds are regulated under the same framework as hedge funds. For example, in Europe, managed futures managers are classified as alternative investment fund managers (AIFMs). AIFMs are regulated under the Alternative Investment Fund Managers Directive (AIFMD). In order to solicit business, a manager of such a fund must register as an AIFM and follow certain regulatory and reporting rules in order to qualify for a European passport to solicit business in the EU. Historically, foreign exchange has been one area of the managed futures industry that has remained largely unregulated. The vast majority of currency trading is conducted in the over-the-counter (OTC), interbank (spot), and forward markets, which are subject to only limited regulation. After the 2008 financial crisis, there has been increased regulatory tightening on all OTC markets.

Title VII of the Dodd-Frank Act in the United States and the European Market Infrastructure Regulation (EMIR) in Europe push for increased transparency and standardization of OTC products. This regulatory push has created an incentive to move many traditionally OTC contracts from bilateral contracts to the multilateral cleared contract structure of futures markets. Many in the industry term this movement from traditional OTC contracts to multilateral cleared contracts the **futurization** of OTC contracts. For the managed futures industry, this means that there is an incentive for growth in futures and a potential increase in the number of tradable futures contracts going forward. The managed futures industry has grown in tandem with the size of futures markets. Exhibit 25.1 plots the number of unique programs in managed futures since 1975. Exhibit 25.2 plots the growth in assets under management in managed futures since 1980. Given these two graphs, it is clear that managed futures programs and assets under management have grown substantially over the past several decades. Futures markets have grown in tandem. To demonstrate this, Exhibit 25.3 plots the growth in volume as a multiple of the volume in 2000 for both established markets (those that traded in 2000) and all markets (established markets plus those added since 2000) versus the growth in assets under management in managed futures. From this graph, it is clear that the managed futures industry is growing in tandem with futures markets.

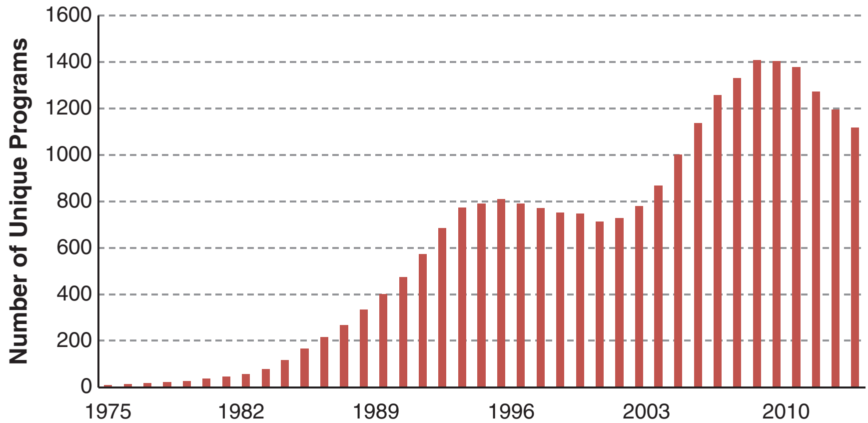


EXHIBIT 25.1 Unique Managed Futures Programs from 1975 to 2013
Source: Greyserman and Kaminski (2014).

25.2 FOUR CORE DIMENSIONS OF MANAGED FUTURES INVESTMENT STRATEGIES

Managed futures strategies can be divided across four core dimensions: data sources, implementation style, strategy focus, and time horizon. Exhibit 25.4 presents a diagram of managed futures strategies.

25.2.1 Data Sources as a Core Managed Futures Dimension

Managed futures strategies are often denoted as either fundamental or technical. Fundamental strategies rely on such data as economic forecasts, supply and demand

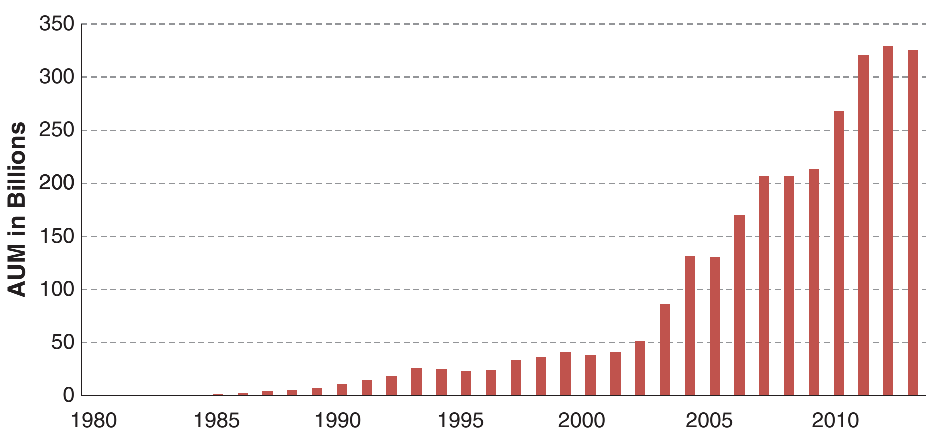


EXHIBIT 25.2 Assets under Management (in $ billion) in Managed Futures from 1980 to 2013
Source: Greyserman and Kaminski (2014).

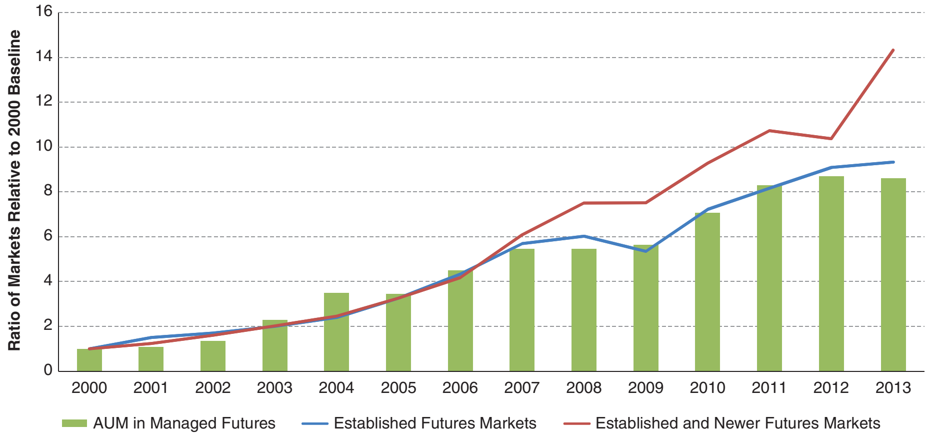


EXHIBIT 25.3 Average Relative Trading Volume for Both Established and Newer Markets and Assets under Management
Source: Greyserman and Kaminski (2014).

estimates, and crop rotation schedules, whereas technical strategies typically analyze historical information such as price and volume. Strategies are also categorized into implementation styles of either systematic or discretionary. Systematic (or quantitative) strategies follow a series of rules to determine entry and exit conditions, position scaling, and position sizes. These strategies rely mostly on the outputs of quantitative models rather than the manager's direct intervention. In terms of turnover, systematic strategies may vary greatly in terms of trade horizon, where trades are held from seconds to months. Over time, systematic strategies have grown in complexity, designed by quant teams that develop the models and automate trading execution. Systematic trading programs also require extensive data capabilities, trading support, and technical support. Systematic implementation is often used in the futures space, given the complexities of futures trading. Trading systems allow a CTA to allocate across

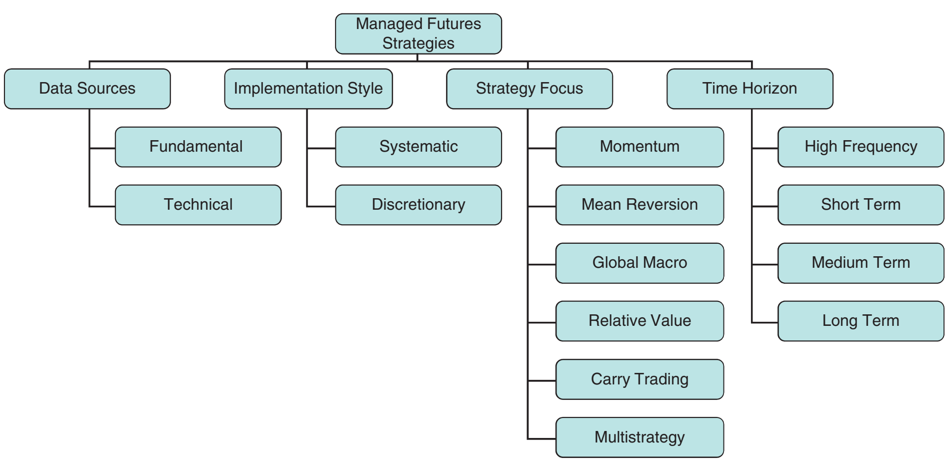


EXHIBIT 25.4 Dimensions of Managed Futures Strategies

many futures markets, maintain collateral, and control and monitor a large number of positions.

25.2.2 Implementation Style as a Core Managed Futures Dimension

Systematic programs are typically more broadly diversified than discretionary traders, both in the number of markets analyzed and in the types of strategies employed. Discretionary managed futures strategies are implemented at the discretion of a CTA manager. These strategies seek to participate opportunistically in market-driven price actions, with the final trading decision being made at the discretion of the fund manager. Since many discretionary managers also use quant models to determine positions, the line between purely discretionary and systematic can sometimes be blurred (Park, Tanrikulu, and Wang 2009).

25.2.3 Strategy Focus as a Core Managed Futures Dimension

Given the range of possible asset classes in futures markets, the range of strategies is relatively broad. Common strategies in the space include momentum, mean reversion, global macro, relative value, carry trading, multistrategy, and many others. In this section, the most common strategies are discussed, with simple examples where applicable.

25.2.3.1 Momentum as a Strategy Focus Momentum is defined as the rate of acceleration of a security's price changes. If prices exhibit momentum, this means that prices are likely to continue in the same direction for some period. For example, if the price of oil is currently $40 a barrel, and it has been moving steadily downward from $50 over the past week, a short-term momentum trader would try to measure the strength of that momentum using simple technical rules. In the case of oil prices moving downward, the trader would most likely take a short exposure to oil in a futures contract.

Momentum is a well-documented phenomenon in traditional asset classes in the body of literature, beginning with Jegadeesh and Titman (1993, 2011), who document momentum in stock returns. Most of the findings in the literature document momentum in the cross section of a particular asset class. For example, cross-sectional momentum in equities would assume that stocks that have under- or outperformed in the past will continue to do so for some period. A cross-sectional momentum strategy would thus go long the outperformers and short the underperformers.

In futures markets, an analogous strategy would be to buy outperforming assets and to sell underperforming assets. For example, if oil prices have been moving downward and cotton prices have been moving upward, a momentum strategy would take short positions in oil futures contracts and long positions in cotton futures contracts. Instead of focusing on one asset class, a momentum strategy in futures markets will be long and short different futures markets across asset classes across time. This aspect of momentum trading is why the approach has been labeled **time-series momentum** in the financial literature. Moskowitz, Ooi, and Pedersen (2012) document the momentum in futures prices over time and coin the term *time-series momentum*. They demonstrate that momentum can be observed across a wide range of time horizons, and profits can be generated by using momentum signals to trade across various asset classes. In the financial industry, momentum trading has long held the moniker of trend following. This is because momentum traders attempt to go long when the price trend is upward and short when the price trend is downward. In simple terms, they are following the prevailing trend in a market. Despite its recent popularity in the financial literature, trend following has been a popular strategy since the 1970s.

In order to implement a trend-following strategy, the trend needs to be measured based on past prices. If a trend-following strategy uses short horizons to measure trends, it is referred to as short-term trend following, while use of long horizons is called long-term trend following. A simple method for measuring a trend (or the strength of a momentum signal) is to use moving averages and compare them to the current price. Consider a short-term momentum strategy on oil. If over the past week oil had an average price of $45 a barrel and the current price is $40 a barrel, the trend is considered downward, and the position will be short oil. If the average price in the past week had been $35 a barrel, the trend would be considered upward, and the position would be long oil.

Although there are many ways to measure a trend across the industry, there are two types of core simple strategies that often provide the basis for trend signals. The first types are moving average strategies and the second are breakout strategies. A moving average is simply a rolling average value of the price. A simple moving average strategy has a long signal when the current price is above the moving average and a short signal when it is below. Exhibit 25.5 plots an example of a simple 20-day moving average for the S&P 500 Index from June 2013 to December 2013. In this example, the trend signal will go long when the price crosses above the 20-day moving average and go short when the price crosses below the 20-day moving average. This example also shows several entry points for long and short signals. A **moving average crossover strategy** uses moving averages across different windows coupled with crossover rules to determine when a trend signals that it is time to take long or short positions. For example, the simple case is based on using a fast moving

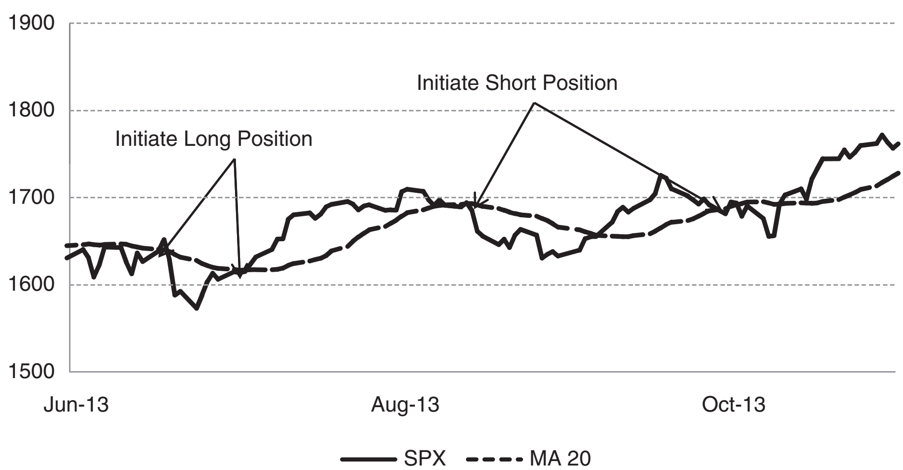


EXHIBIT 25.5 An Example of a 20-Day Moving Average Signal on the S&P 500 Index from June 2013 to December 2013

average and a slow moving average. More explicitly, a 150-day moving average of the price could be the slow moving average (i.e., considers a longer window), and a 50-day moving average of the price could be the fast moving average (i.e., considers a shorter window). A moving average crossover strategy has a long trend signal when the fast moving average is greater than the slow one, and a short trend signal when the fast moving average is less than the slow one. One key issue with a simple application of this strategy is that trend signals based on moving averages may change sign often. This can result in unwanted trading and the associated trading costs. At the core, trend signals based on moving averages follow the trends as they evolve.

A breakout strategy creates a positive or a negative trend signal when the price breaks out of a range of values. This range of values is often referred to as **resistance levels** and **support levels**. Resistance and support levels can be defined using many different techniques, including past prices and trading range. However, very often these levels are simply the high and low prices over a given look-back window. For example, the highest price of a commodity over the past 10 days may represent the resistance level, while the lowest price of the same commodity over the same period may represent the support level. A breakout strategy has a positive trend signal when the price goes above the resistance level and a negative trend signal when the price goes below the support level. Trading signals based on breakout trend signals are designed so that the position exits when either an opposite breakout trend signal occurs or a **trailing stop** is reached. A trailing stop is a stopping rule that depends on the recent path of the price such that the stop "trails" the current price. For example, a trailing stop price may be set to equal 2% below its most recent high. If the current price is $100, then the position will be exited if the price goes below $98. If the price increases to $110, then the trailing stop price will increase to $107.80.

Both a moving average strategy and a breakout strategy define a set of rules to create trading signals. A **trading signal** is a signal that defines the position in a particular market long or short. These signals can be raw trend signals, filtered versions of trend signals, or a combination of both. In practice, trading signals often aggregate trend signals from many different methods and across different sets of parameters. For example, a trading signal might be the average of several moving average signals from different time horizons and a range of look-back windows. Trend-following strategies measure the strength of trends across futures markets. Futures trading strategies are built by allocating risk across positions in futures markets. The final section in this chapter discusses the specifics of how futures positions are determined and how futures portfolios are constructed.

25.2.3.2 Global Macro as a Strategy Focus Global macro managed futures strategies use fundamental information to determine long and short allocations across the global range of futures markets. (Global macro is further discussed in Chapter 28 of this book.) Global macro futures strategies will take positions in foreign exchange, commodities, domestic and international equity, and fixed-income markets based on fundamental analysis. A global macro strategy can develop economic models that attempt to explain how the global economy will react to various changes in economic regimes. For example, if the U.S. dollar appreciates, this improves the purchasing power of the U.S. dollar while decreasing the attractiveness of U.S. exports. A global macro manager can construct an economic model that attempts to predict both U.S. dollar exchange rates and the economic impact on

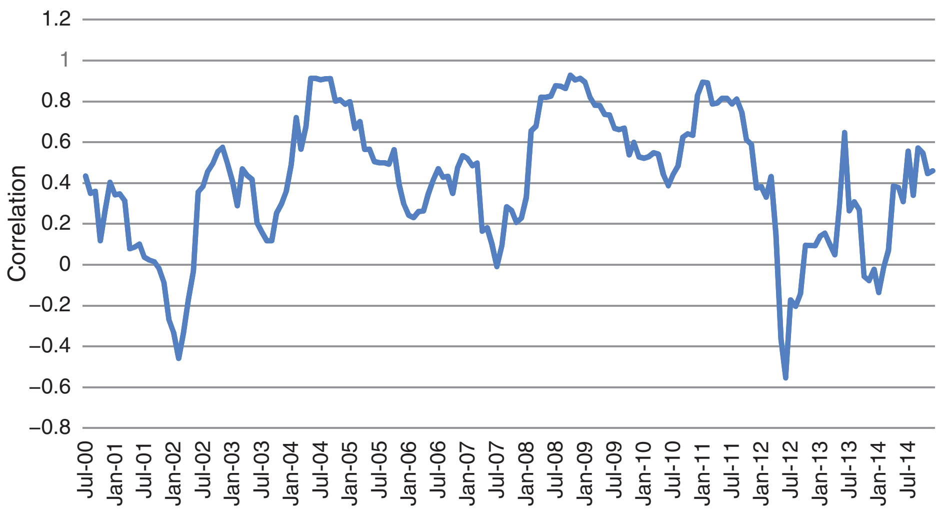


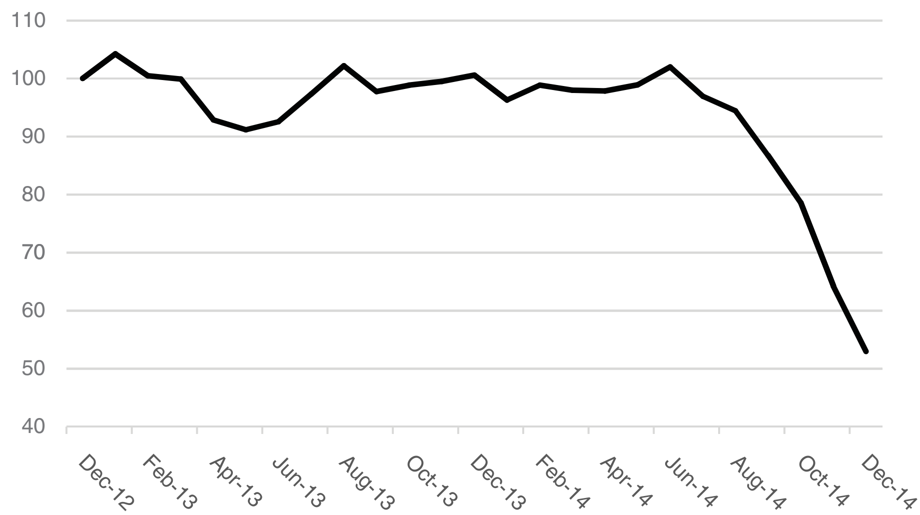
EXHIBIT 25.6 Rolling 12-Month Correlation: Systematic and Discretionary CTA Strategies
Source: Bloomberg and authors' calculations.

commodities, global stock markets, and short- and long-term rates. Another macro strategy may use fundamental inputs and data to gauge the change in the demand for oil. If a global macro manager's model predicts an increase in demand, the manager will go long oil; if demand is predicted to decrease, the manager will short oil.

Global macro managers and trend-following strategies use different inputs to determine where global trends will occur. Because of this, the performance of discretionary versus systematic managed futures can vary. Exhibit 25.6 charts the rolling 12-month correlation of the CTA Barclay Systematic and Discretionary Indices from 1997 through 2014. The 12-month correlation turned negative twice during the period: following September 11, 2001, and again in 2012. Another key difference between the two is that a fundamentally based strategy is more likely to enter a global trend earlier than a trend-following strategy will. This is because trend-following strategies need to measure momentum in prices to begin a position. Although global macro strategies often get in earlier, the market may not agree with the fundamentals for long periods of time, causing difficulty for global macro strategies.

A good example would be oil futures prices in 2014 (see Exhibit 25.7). In 2014, the front month oil futures contract price plummeted by about 60%. The momentum in this price series was clearly pronounced, and many trend-following strategies picked up the momentum signal and went short oil. However, this particular event was not easily predicted by global macro models, with many market participants feeling that it was unlikely for oil to drop so aggressively. In this scenario, a trend-following strategy would have profited from going short, while many fundamental strategies would have continued to be long oil.

25.2.3.3 Relative Value as a Strategy Focus Relative value strategies focus on finding relative mispricing between different assets across markets or across time. There are many ways to implement a relative value strategy in futures. Common



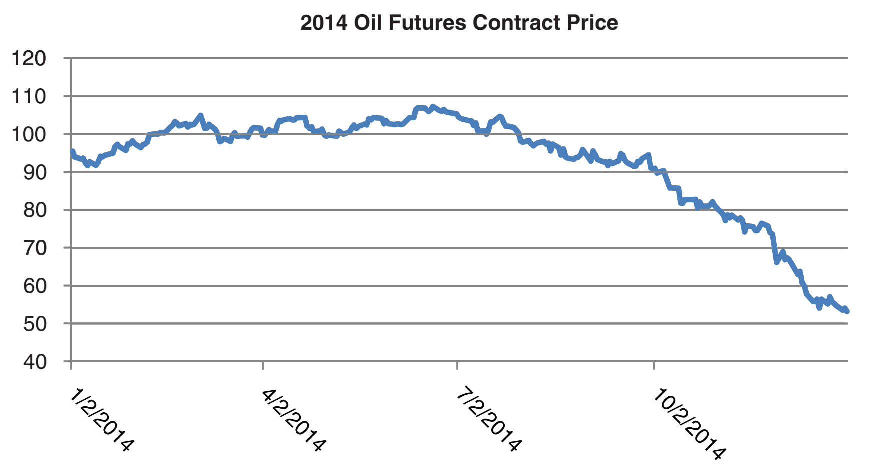


EXHIBIT 25.7 S&P GSCI Brent Crude Spot Price, December 2012 to December 2014 (top); 2014 Oil Futures Contract Price (bottom)

examples include calendar spreads, crack spreads, crush spreads, and cross-asset relative value spreads. For example, calendar spread strategies take long and short positions across different maturities of futures contracts.

25.2.3.4 Other Strategy Foci Mean-reversion and countertrend strategies focus on prices reverting to the mean in the short term. Mean-reversion strategies are typically implemented in days to weeks, whereas trend following is often implemented in months. A strategy is considered contrarian if it trades against the prevalent trend. Mean-reversion strategies often take contrarian positions. A contrarian strategy would have the opposite sign of a trend-following strategy.

Carry strategies are designed to take advantage of differences in the carry of various commodities. In general, the carry of an asset is the return obtained from holding it (if positive) or the cost of holding it (if negative). For instance, commodities are usually negative carry assets, as they incur storage costs and may suffer from depreciation; however, appropriately hedged commodities can be positive carry assets if the futures market is willing to pay a sufficient premium for future delivery. The carry trade, which involves long positions in fixed-income futures of high interest rate countries, and short positions in fixed-income futures of low interest rate countries, is explained in further detail in Chapter 28. **Multistrategy CTAs** combine a variety of strategy focuses to provide a diversified set of potential return sources and risk-reward profiles.

25.2.4 Time Horizon as a Core Managed Futures Dimension

The final aspect that differentiates managed futures strategies is time horizon. Time horizons can range from high frequency to long term. Trading speed is often measured by the average holding period for each trade. High-frequency trading is reserved for a special group of strategies that do not attempt to benefit from trends in prices and traditionally have not been classified as managed futures strategies. Short-term CTA trading strategies are often classified as intraday to one month, with an average holding period of around 10 days or less. Medium term can be one month to six months. Long term can be greater than six months. Mean reversion and countertrend tend to be shorter term, whereas trend following is often medium to long term. It is very common for systematic funds to combine a large range of approaches, from short term to long term.

A core issue in dynamic futures trading strategies is transaction costs, trading capacity, and slippage. Transaction costs are incurred on a per trade basis. As a result, transaction costs are very important for short-term strategies that may require frequent transactions. Medium- to long-term strategies focus on longer-term effects. As a result, they do not adjust their positions as quickly or as often as do shorter-term programs. Trading capacity is important for strategies that take large positions relative to open interest. If a strategy requires large short-term changes in positions, the strategy may move the market against the strategy. Since CTAs manage accounts on behalf of clients, they often trade multiple accounts simultaneously. This could lead to slippage. Slippage occurs when actual performance deviates or "slips away" from the expected trading results using the computer's signal.

25.3 FOUNDATIONS OF MANAGED FUTURES

Dynamic strategies attempt to capture opportunities in financial markets. The concept of market efficiency has direct implications for understanding when investment strategies can capture opportunities. The classic efficient market hypothesis (EMH) is a framework for understanding how market prices behave. In its strongest form, it declares that all relevant information is incorporated into prices. Even in its weakest form, the EMH implies that historical price information cannot be used to make profitable investments. The majority of managed futures strategies are systematic, including the trend-following approach, which is the most common strategy. Managed futures managers typically trade the most liquid and, seemingly, the most efficient markets. This construction should, by all accounts, be in conflict with the traditional efficient market hypothesis. Given this, it is not surprising that the efficient market hypothesis may not be the right framework for understanding managed futures strategies. Instead, an alternative theory for explaining market behavior, the adaptive markets hypothesis (AMH), may be necessary.

25.3.1 The Adaptive Markets Hypothesis

As first proposed by Lo (2004), the **adaptive markets hypothesis (AMH)** is an approach to understanding how markets evolve, how opportunities occur, and how market players succeed or fail based on principles of evolutionary biology. According to the AMH, concepts central to evolutionary biology govern market dynamics via the forces of competition, mutation, reproduction, and natural selection:

> *Prices reflect as much information as dictated by the combination of environmental conditions and the number and nature of "species" in the economy, or . . . ecology. Species are defined as distinct groups of market participants: e.g., pension funds, retail investors, hedge funds. (Lo 2004)*

According to the AMH, profit opportunities exist when more resources are present and competition is lower. Through the mechanism of natural selection, as competition increases, the players who have competitive advantage over others survive and adapt. Those who are not able to adapt disappear, reducing competition and starting the evolutionary cycle all over again. An excellent example of this phenomenon at work is the waxing and waning of hedge fund styles.

25.3.2 Four Practical Implications of the Adaptive Markets Hypothesis

According to Lo (2012), some of the important practical implications of an adaptive view on markets include these four:

1. *Time-varying risk premiums.* The trade-off between risk and return is not stable over time (risk premiums vary over time). In addition, changes in risk premiums could be predicted based on technical and fundamental variables.
2. *Market efficiency is a relative concept.* Market efficiency should be measured and discussed in relative terms as opposed to absolute terms. Market efficiency is a continuum; it is not simply efficient or inefficient, and a market displays varying degrees of efficiency at different points in time and for different market participants.
3. *Adaptation for success and survival.* It is necessary to use adaptable investment approaches to handle changes in the market environment. Opportunities are not always found in the same place; therefore, trading strategies must be altered as the economic environment evolves.
4. *The inevitable degradation of alpha.* With time, what was once alpha becomes, due to innovation and competition, beta. Persistent alpha opportunities are not possible; however, fleeting alpha opportunities may be possible.

Based on the implications of the adaptive markets hypothesis, the use of systematic futures strategies can be put into a different perspective. According to the first implication, risk premiums are time varying, which implies that strategies that allocate dynamically across time may be able to capture these risk premiums. Second, because market efficiency is relative across time, there may be periods when markets may be less efficient, as the "market ecology" adapts and adjusts. This suggests that strategies that focus on exploiting market efficiencies can find opportunities from time to time. Third, adaptive strategies can provide opportunities during periods when other strategies may suffer. The financial crisis is a good example of a situation during which this point became relevant. Fourth, consistent alpha opportunities should degrade over time, but at certain moments it may be possible for highly adaptive strategies to find opportunities.

25.3.3 Market Divergence, Dislocation, and Momentum

According to evolutionary biology, *divergence* is defined as the evolutionary tendency or process by which animals or plants that are descended from a common ancestor evolve into different forms when living under different conditions. In financial markets, **divergence** can be defined as the evolutionary tendency or process by which the market ecology evolves in response to changes and structural shifts in the economy. If financial markets were perfectly efficient and they evolve in a frictionless way, divergence in market prices would not lead to the emergence of opportunities with abnormally high risk-adjusted returns.

The AMH states that markets adapt and adjust to changes in the financial environment. As the environment changes, competition spikes among market players, and shocks disrupt market processes. Prices reflect the current environment and the level of competition among market participants. Similar to a biological system, shocks move the market to a new equilibrium condition. This process of change and adaptation leads to divergence. That is, a changing economic environment and the reaction of market participants to these changes move prices away from the no-arbitrage level, creating temporary profit opportunities. This process is not simultaneous and may evolve slowly until prices converge to the level of no arbitrage. Of course, this process never stops, as the environment is constantly changing; therefore, prices are always in a state of divergence and convergence. However, these conditions take place at different corners of the market, which means a traditional static strategy would not be able to take advantage of opportunities that may arise.

Similar to Moskowitz, Ooi, and Pedersen (2012), most empirical studies of momentum simply suggest a long list of causes for momentum. These include a wide range of explanations, from behavioral bias to market frictions. From an adaptive markets perspective, momentum is simply caused by slow or prolonged periods of market divergence. Just as in evolutionary biology, financial systems are made up of people and computers run by people. Pasquariello (2014) presents empirical evidence for the existence of market dislocation, which is somewhat similar to divergence. Exhibit 25.8 presents a summary of the definitions of divergence, dislocation, and momentum. Both the biological definition and the corresponding financial interpretation of each term are presented, to show the parallel between the two as well as the connection to evolutionary biology.

EXHIBIT 25.8 Divergence, Dislocation, and Momentum

Term	Biological Definition	Financial Interpretation	Causes
Divergence	The evolutionary tendency or process by which animals or plants that are descended from a common ancestor evolve into different forms when living under different conditions	The process through which market participants and groups of market species evolve and adapt to new market conditions	Shifts in risk appetite, changes in supply and demand, behavioral bias, sentiment, crises, market frictions, systemic risks
Dislocation	The act of displacing or the state of being displaced; disruption from the norm or steady state	Prices moving away from no-arbitrage relationships	Slow or prolonged divergence
Momentum	Force or speed of movement for a physical object or a course of events	When prices moving in one direction continue to do so for some period of time	Slow or prolonged divergence

25.3.4 Measuring Market Divergence

Similar to how volatility can be measured by standard deviation, market divergence can be measured empirically. In Exhibit 25.9, several price trends are plotted, first with a clean price trend and then with low noise and high noise in the price series. In this exhibit, 100-day signal-to-noise ratios (SNRs) are calculated for each 100-day example. The noisy linear price series has a return volatility of 10%, and the very noisy linear price series has a volatility of 40%. On the left-hand side, a 100-day trend is plotted. For comparison, a shorter and steeper 20-day trend, taken from the 100-day period, is also plotted. The level for both trends is the same: a $10 increase in price.

To measure the level of divergence for a particular price series, it is necessary to examine the signal-to-noise ratio. The **signal-to-noise ratio** is the ratio of the overall trend to a series of price changes during the same period. Signal-to-noise ratio can be viewed as the ratio of the magnitude of a trend to the volatility around that trend. If, for instance, the uptrend in the market is achieved through a number of positive and negative changes in prices, then the signal-to-noise ratio is small. On the other hand, if the uptrend is the product of a number of positive changes in the price, then the signal-to-noise ratio is high. Examining the formula that calculates the signal-to-noise ratio will help in understanding this concept. For any specific day, at time t, the signal-to-noise ratio (SNR_t) for a particular price series with look-back period (n) can be calculated mathematically using the following formula:

$$SNR_t(n) = \frac{|P_t - P_{t-n}|}{\sum_{i=0}^{n-1} |P_{t-i} - P_{t-i+1}|} \tag{25.1}$$

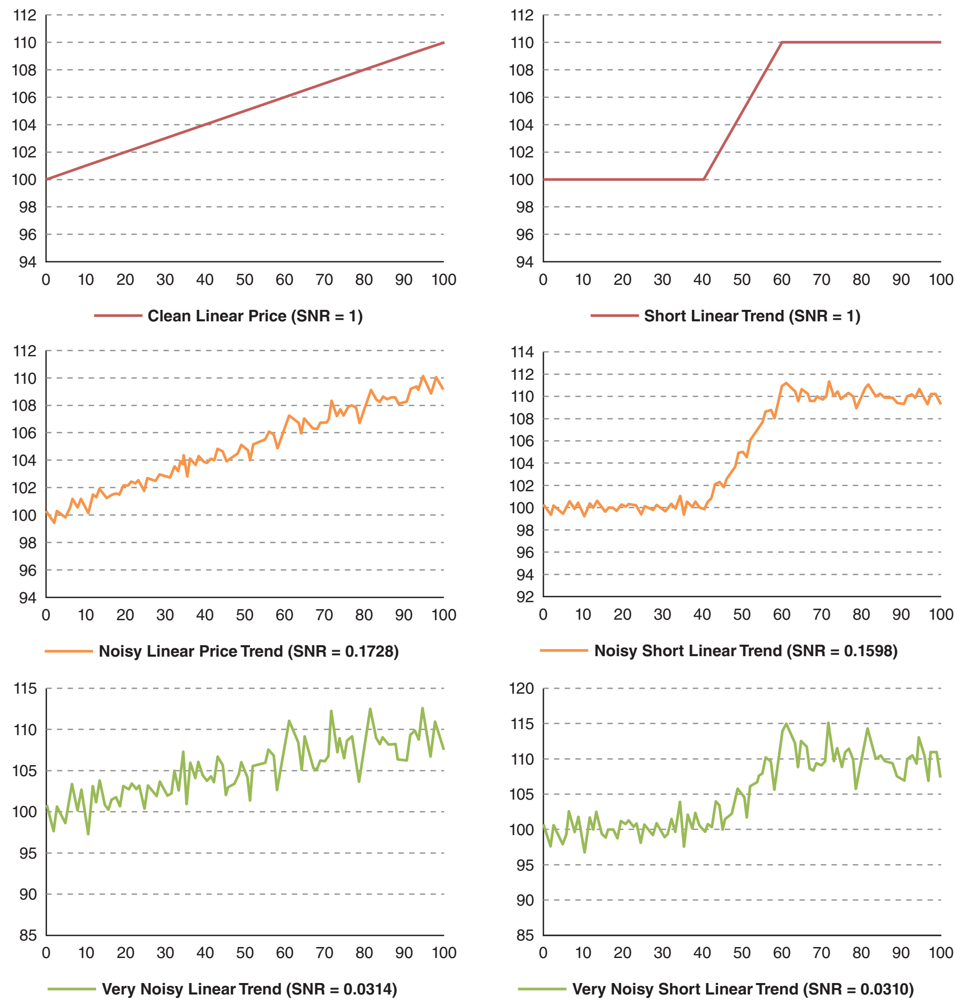


EXHIBIT 25.9 Examples of Price Trends with and without Noise

where P_t is the price at time t, and n is the **signal observation period** (i.e., the look-back window for the calculation of the signal). Suppose the numerator of the equation is positive (1), which means that during the past n days, the price has increased (i.e., there is a positive trend). Further, suppose this price increase took place gradually over the past n days, with the price rising a little each day. In this case, the ratio will be equal to 1, indicating that there was no noise. On the other hand, suppose that the price rises in an uneven manner, increasing on some days and decreasing on other days. Eventually, the net increase equals the numerator, but because the denominator uses the absolute values of the changes, the ratio will be less than 1, indicating that there was some noise in the trend. In fact, if the ratio is too small, a CTA may conclude that there was no trend at all and that the price action was driven entirely by noise.

For medium- to long-term trend followers, the length of the look-back window is typically chosen at roughly 100 days.

APPLICATION 25.3.4a

For a numerical example, if $n = 5$, there are five days in the signal observation period, and if the price series is 50, 51, 52, 50, and 53, then the SNR would be found using Equation 25.1 as $|53 - 50|$ divided by $|51 - 50| + |52 - 51| + |50 - 52| + |53 - 50|$, which is 0.43. This indicates that the trend is reasonably strong. If the final price were 51 instead of 53, the SNR would be 0.2, indicating a weaker trend.

Returning to Exhibit 25.9, the 100-day SNR is given for each price series. The perfect price trends have a signal-to-noise ratio of 1. A higher SNR indicates a higher quality of the trend, or a higher price divergence for that individual market. Notice that as noise (or volatility) enters into price trends, SNR becomes lower, as the trend is much harder to discern. For example, the SNRs for the noisy and very noisy linear price trends are 0.1728 and 0.0314, respectively. The noisy linear price trend would be interesting for a trend follower, but the very noisy linear price trend would be deemed too volatile. This simple example explains how trend followers have a complex relationship with volatility.

Adding volatility to a price series reduces the attractiveness of a price trend. On the other hand, when new trends emerge in prices, price series exhibit higher volatility because they are moving away from the average values established during the trendless period. This is sometimes called directional volatility. The right-hand side of Exhibit 25.9 plots a shorter trend, with the same increase in the price over a 100-day period. In terms of level, these two price moves are the same, yet the second price trend would be less desirable from a 100-day perspective. Notice that the SNR for the noisy linear price trend is 0.1728 versus 0.1598 for the short-term price trend. The same noise series is used in this example to make the SNRs comparable. When the annualized volatility is 40%, both the very noisy linear and short-term price trends are relatively useless.

The SNR can be calculated for each individual market. The next step is to look at the aggregate level of divergence in markets at the portfolio level. For each individual market i, for simplicity the total level of market divergence (or "trendiness") can be calculated as the average signal-to-noise ratio. This quantity can be defined as the **market divergence index (MDI)** for a given signal observation period n. In particular, when there are M markets, MDI is calculated as follows:

$$MDI_t(n) = \frac{1}{M}\sum_{i=1}^{M} SNR_t^i(n) \tag{25.2}$$

where SNR_t^i is the signal-to-noise ratio for an individual market i, and n is the signal observation window. The MDI is a simple aggregate measure of divergence in prices, taking into account the level of volatility (or noise) in the price series. When the MDI is higher, this corresponds to a market environment with stronger trends across markets in a portfolio.

APPLICATION 25.3.4b

Suppose the S&P 500 Index, Japanese government bonds (JGBs), corn, wheat, crude oil, and Treasury bonds have 100-day SNRs of 0.3, 0.2, 0.3, 0.1, 0.4, and 0.3, respectively. Crude oil has the largest SNR, which means that the trend signal is the highest for crude oil over the past 100 days. Wheat prices have the smallest SNR, which means that the trend signal is the lowest. Using Equation 25.2, the MDI for all included markets would be the average SNR. In this case, MDI = 0.267. This figure can be compared to SNRs from previous periods to determine whether markets are trending. The MDI is a useful measure for diversified CTAs.

According to the AMH, since futures markets are highly competitive, measured divergence in prices should be very low, with occasional periods during which market prices deviate from a random walk. Exhibit 25.10 plots a histogram of measured market divergence across a large set of futures markets from 2001 to 2013, where the horizontal axis is the MDI value, and the vertical axis is the number of samples in each bin of the MDI value. Points with high market divergence, on the right side of the distribution, coincide with periods of financial crisis, when markets may be less efficient. Managed futures strategies could use this construction to profit from periods of market divergence. Exhibit 25.11 plots the conditional performance of a representative trend-following strategy across a large set of futures markets as a

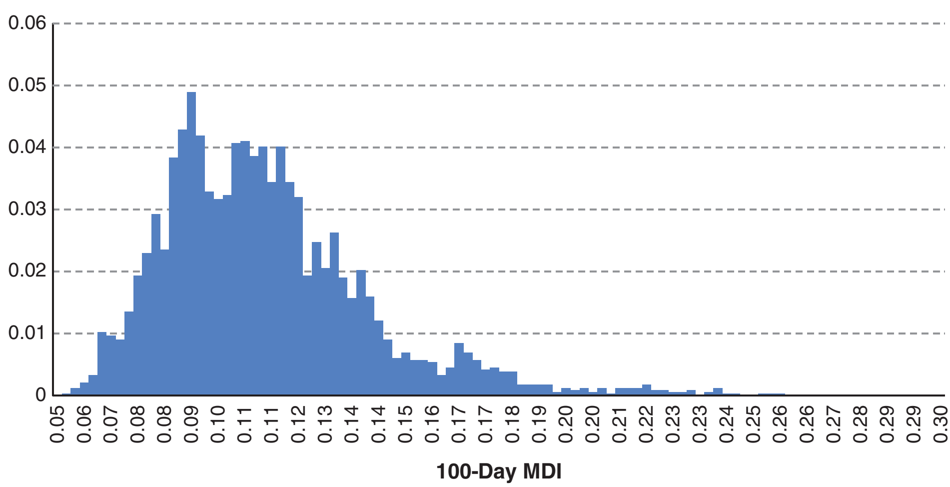


EXHIBIT 25.10 Histogram for the MDI from 2001 to 2013 for a Portfolio Consisting of Commodities, Equity Indexes, Fixed Income, and Currencies

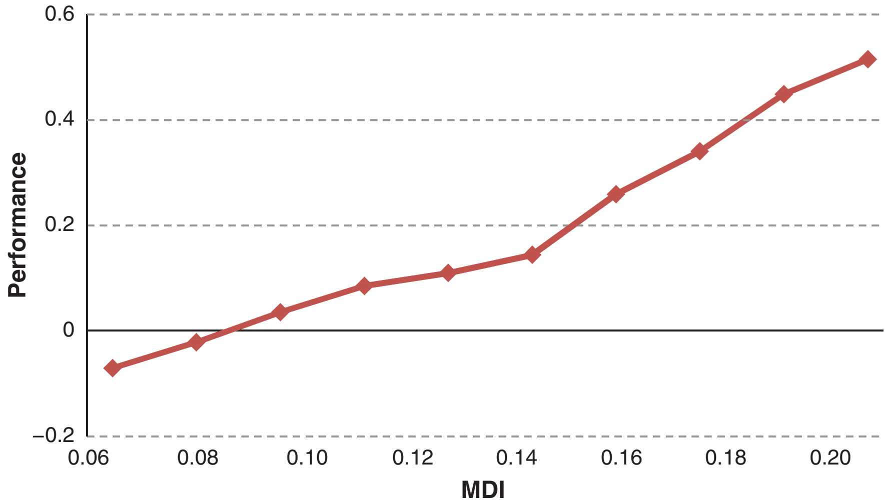


EXHIBIT 25.11 The Conditional Performance of the 100-Day Rolling Return for a Representative Pure Trend-Following Program as a Function of the MDI

function of market divergence (100-day MDI). Here, the horizontal axis is the MDI value, and the vertical axis is the conditional mean of the 100-day return for each corresponding MDI value. This figure demonstrates how a systematic strategy can take advantage of opportunities when measured market divergence has been high. The correlation between the 100-day MDI and the performance of a representative trend-following program is 0.74 in the sample period.

25.3.5 Crisis Alpha

Given that managed futures strategies tend to perform well when markets diverge, they are a natural complement to many traditional strategies that may suffer during periods of market stress. The financial crisis provides an extreme yet relevant example. Times of market crisis, for both behavioral and institutional reasons, represent times when market participants become synchronized in their actions, creating trends in markets. It is only the most adaptable market players—a select few—who can take advantage of these "crisis alpha" opportunities.

When equity markets go down, the vast majority of investors may experience losses. At these times, investor decisions may be influenced by behavioral biases and emotionally based decision-making. When coupled with the widespread use of institutionalized drawdown and leverage, along with risk limits triggered by losses, increased volatility, and increased correlation, these losses will drive large groups of investors into action. When large groups of investors are forced into action, liquidity disappears, credit issues come to the forefront, fundamental valuation becomes less relevant, and persistent trends occur across all markets while investors fervently attempt to exit positions. Opportunistic systematic strategies that are void of long equity bias, highly adaptive, and liquid can adapt to capture crisis alpha opportunities.

EXHIBIT 25.12 The Characteristics of Trend Following and Their Implications during Periods of Equity Market Crisis

Characteristics of Trend Following	Implications during Equity Market Crisis
Highly liquid, adaptable strategies based exclusively in futures with minimal credit exposure	Less susceptible to the illiquidity and credit traps that most investors experience during equity market crises
Systematic trading strategies, lack of long equity bias	Less susceptible to behavioral biases and emotionally based decision-making triggered by experiencing losses
Active across a wide range of asset classes in futures	Poised to profit from trends across a wide range of asset classes

It is important to note that systematic strategies do not time the onset of an extreme event, like an equity market crisis, but may profit from opportunities in the aftermath of a crisis in currencies, bonds, short rates, equities, and commodities. Using systematic trend following as an example, its characteristics and their implications during an equity market crisis are summarized in Exhibit 25.12.

25.4 BENEFITS OF CTAs

A more direct way to examine the attractiveness of a managed futures investment is to analyze the performance of managed futures traders.

25.4.1 Research Regarding the Benefits of CTAs

In 1983, John Lintner presented one of the first academic papers on the topic of CTAs and their benefits. His analysis was designed to examine the risk-return characteristics of managed futures accounts or funds. In this study, Lintner concluded that "the combined portfolios of stocks (or stocks and bonds) after including judicious investments . . . in managed futures accounts (or funds) show substantially less risk at every possible level of expected return than portfolios of stocks (or stocks and bonds) alone."

Lintner's work provided an initial academic basis for investing in managed futures. Other early studies that followed his research both challenged and supported his results. A series of studies by Elton, Gruber, and Rentzler (1987, 1989, 1990), known as the EGR studies, examined public commodity pools and, unlike Lintner, found little evidence of the benefits of managed futures. However, other analyses of managed futures supported their inclusion in investment portfolios. Some of these later analyses attempted to address data issues in the EGR studies. The EGR studies looked at public commodity pools, known to have been a very expensive way to invest in managed futures. Later analyses directly examined the returns of managed futures traders and found evidence that, on average, managed futures provide attractive risk-adjusted returns, especially if the performance is measured in the context of a diversified portfolio of stocks and bonds.

Unlike previous studies that examined the benefits of CTAs either as a stand-alone investment or in the context of portfolios consisting of traditional asset classes, Kat (2002) examines the possible role of managed futures in portfolios of stocks, bonds, and hedge funds. He finds that allocating to managed futures allows investors to achieve a very substantial degree of overall risk reduction at limited costs in terms of lower returns or skewness. Apart from their lower expected returns, managed futures appear to be more effective diversifiers than hedge funds. The paper concludes that adding managed futures to a portfolio of stocks and bonds will reduce that portfolio's standard deviation more and quicker than hedge funds will, and without the undesirable side effects in terms of lower skewness and higher kurtosis. Finally, Kat observes that overall portfolio standard deviation can be further reduced by combining both hedge funds and managed futures with stocks and bonds. Again, it is worth repeating that these results are, to some degree, time dependent and can change dramatically over short periods of time.

Some studies have shown that systematic trend-following strategies tend to outperform discretionary strategies on a risk-adjusted basis (Aldridge 2009). These studies show that on the basis of absolute monthly returns, systematic funds outperform discretionary funds whenever the relevant markets are falling. When markets are rising, however, discretionary funds tend to deliver higher absolute returns than do systematic funds. Across a variety of metrics, systematic funds perform better than discretionary funds. In particular, on an ex post basis, systematic funds produce less extreme drawdowns, higher Sharpe ratios, and higher Jensen's alpha (measured against traditional asset classes). Furthermore, systematic funds exhibit lower skewness and kurtosis than do discretionary funds. Results suggest that much of systematic funds' outperformance comes from a better ability to manage extreme events, performing better than discretionary funds in crisis conditions.

Managed futures managers and trend followers, in particular, are supposed to be able to time their chosen markets (e.g., currencies or commodities). A study by Kazemi and Li (2009) investigated the return and volatility market-timing ability of CTAs and examined whether there is a difference in market-timing abilities between systematic and discretionary traders. For this purpose, a set of risk factors was developed based on returns from the most heavily traded futures contracts. The study made two conclusions:

1. CTAs exhibit both return and volatility market-timing ability in markets that they have declared to be their focus, most notably for currencies, interest rates, and commodities. However, CTAs display negative return timing in equity markets.
2. Systematic traders are generally better market timers than are discretionary traders. Systematic traders show timing ability for currency futures and physical product (corn, crude oil, natural gas, gold) futures.

25.4.2 Sources of Return for CTAs

The sources of return to managed futures are essentially different from those related to traditional stocks, bonds, or even hedge funds. For instance, futures, swaps, and forward contracts can provide direct exposure to underlying financial and commodity markets but often with greater liquidity and less market impact. Futures and options allow traders to take short positions without the need to borrow the securities from other investors. This allows traders to actively allocate assets between long and short positions within the futures/options market-trading complex. In addition, options traders may also directly trade market/security characteristics, such as price volatility, that underlie the contract. The unique return opportunities of managed futures may also stem from the global nature of futures contracts available for trading and the broader range of trading strategies.

It is important to note that many managed futures strategies trade primarily in futures markets, which in aggregate are zero-sum games. This means that for each amount that is gained by a group of traders, there is an equal amount that is lost by another group of traders. (In traditional equity and fixed-income markets, investor gains are not offset by losses by other investors.) The implication is that if, as an asset class, CTAs provide positive returns on a consistent basis, then other investors must be earning negative returns on a consistent basis. This may appear to be implausible. Futures markets are not subject to this implication of a zero-sum game if one considers that many participants have positions in cash markets and that losses in futures markets may be offset by gains in cash markets. In other words, some spot market participants may be willing to lose money in futures markets because the loss is offset by gains in other parts of their operations. These gains could be in the form of higher income or lower risk. The classic case of a corn farmer who hedges and is willing to experience a small loss on the futures contract in exchange for avoiding exposure to fluctuations in future corn prices highlights this scenario. Without a hedge, this farmer could lose her entire business if the price of corn unexpectedly declines. Thus, she would be willing to suffer a small loss in the futures market to avoid the risk of ruin.

Since in broader terms futures are not a net zero-sum game, managed futures strategies can earn a positive rate of return if they provide a service or benefit to those market participants who are willing to accept small losses in futures markets while experiencing higher profits, lower losses, or lower risks in other markets. Managed futures managers may fulfill the following functions: (1) allow other participants to hedge a position and therefore reduce their risks, (2) provide liquidity so that other participants satisfy their need for liquidity, and (3) take offsetting positions for rebalancing and other demands from other market participants. These activities allow CTAs to earn a premium, which is highly time varying. This means that the premium is not earned by always taking the same long or short position in a given market; rather, the premium is earned by taking both long and short positions during different periods. Hedging and liquidity provision is an important aspect of the benefits provided to other market participants. Whereas futures contracts are used by some participants to hedge their risks, spot markets serve a different purpose. Therefore, the same level of premiums may not be earned by a CTA if the trading strategy involves the spot market.

One simple way to examine this effect is to compare the performance of a simple trend-following strategy on both spot and futures prices. Both spot and futures prices are available at the same time, but they represent different things. Spot prices represent the price to purchase or sell something at the current time. Futures prices represent the price to purchase or sell something in the future. Futures contracts allow for risk transfer and hedging. Exhibit 25.13 lists the Sharpe ratio, monthly return, and monthly risk for the performance of a representative trend-following strategy

EXHIBIT 25.13 The Total Futures Price and Spot Price Only Performance for a Representative Pure Trend-Following System, 1980–2013

		Sharpe Ratio	Monthly Return (%)	Monthly Risk (%)
All	Futures	0.74	1.01	4.65
	Spot	0.46	0.63	4.64
Fixed Income	Futures	0.51	1.45	9.64
	Spot	0.21	0.60	9.63
Short-Term Interest Rates	Futures	1.12	2.95	8.90
	Spot	0.98	2.57	8.88
Equity	Futures	0.08	0.20	8.08
	Spot	0.09	0.21	8.08
Commodity	Futures	0.73	1.00	4.60
	Spot	0.35	0.47	4.59
Currency	Futures	0.39	0.93	8.10
	Spot	0.28	0.66	8.10

Source: Greyserman and Kaminski (2014).

on both the spot price series and the futures price series across all asset classes from 1980 to 2013. The Sharpe ratio for trend following on spot prices is 0.46, whereas the Sharpe ratio on futures prices is 0.74. This demonstrates how futures prices may tend to offer more hedging benefits to other market participants, leading to profit opportunities for CTAs. Another interesting aspect of Exhibit 25.13 is the variation in performance across asset classes. The outperformance of the futures prices is most pronounced in fixed-income and commodity markets. This is consistent with the hedging of long-term fixed-income and commodity risk. The performance of equity futures prices is roughly the same as direct investments in equities. This suggests that there may be fewer hedging or liquidity provision opportunities in equities for managed futures.

25.4.3 Eight Core Benefits of CTAs for Investors

CTAs provide a number of benefits in terms of risk-return trade-offs to investors, eight of which are introduced here. For a discussion of these benefits, see Schneeweis (2009) and Burghardt and Walls (2011).

Diversification. Managed futures constitute an alternative asset class that has achieved strong performance in both up and down equity, commodity, and currency markets, and has exhibited low correlation to traditional asset classes, such as stocks, bonds, cash, and real estate. Managed futures, when used in conjunction with traditional asset classes, may reduce risk while potentially increasing portfolio returns.

Performance. Historically, managed futures have provided risk-return profiles comparable to those of many traditional asset classes and superior to those offered by long-only investments in commodities. For example, the historical Sharpe ratio of a diversified portfolio of managed futures could be four times higher than that of a long-only portfolio of commodities.

Access to multiple markets. There are more than 150 liquid futures products across the globe, including stock indexes, currencies, interest rates, fixed income, energies, metals, and agricultural products. CTAs are able to take advantage of potential opportunities in various asset classes in many geographical locations. The fundamental law of active management states that the information ratio of an investment increases as the breadth of the investment strategy increases (holding other variables constant). This means that CTAs have the potential to provide performance with a superior risk-return profile.

Transparency. Futures prices are determined competitively and are marked to market daily. The fact that futures prices tend to be determined in single-price discovery markets in which everyone can see the limit order book and in which the settlement prices are, in most cases, tradable makes them more accurate and more reliable than prices determined in nearly any other market. The prices used to mark portfolios to market are not stale. There are no dark pools of liquidity, like those found in equity markets. There are no interpolation methods similar to those of some bond markets, in which only a handful of bonds actually trade on any given day. Moreover, there are no models needed to determine the value of structured securities. As a result, the returns experienced are real and have not been smoothed.

Liquidity. Liquidity has already been mentioned, but only in the context of liquidating positions and extracting cash. In fact, transaction costs in futures are lower than in their underlying cash markets. As a result, the benefits of the kind of active management and trading that CTAs implement are available with less drag from market impact than one would incur with the same type of trading in underlying markets.

Size. As an investment alternative, managed futures have been available since the 1970s and experienced significant growth over the past several decades. According to the National Futures Association and Barclay Hedge (www.nfa.futures.org and www.barclayhedge.com), as of 2015 the size of the market was estimated at about \$330 billion. This means the market has reached a level at which it can accommodate allocations from institutional investors.

No withholding taxes. In a number of the world's stock and bond markets, foreign investors are taxed more heavily than are domestic investors. With futures, all of the tax benefits that accrue to domestic investors can be passed through to those who use futures in the form of simple cash/futures arbitrage.

Very low foreign exchange risk. Futures on foreign assets or commodities have little exposure to foreign exchange risk. A futures contract has no net liquidating value. As a result, a long position in a European equity index futures contract has no exposure to the change in the price of the euro, whereas an investment in European equities exposes the investor not only to changes in the price of European stocks but to changes in the price of the euro as well. In the case of futures, the investor's currency risk is limited to the comparatively small amounts of margin that must be posted at exchanges around the world and to any realized profit or loss that has not yet been converted back into the investor's home currency.

To understand this benefit, note that a position in a futures contract is similar to a long position in the same asset in the cash market, where the position is financed through borrowing. This means a futures position in a foreign-currency-denominated asset is similar to a cash position in the same asset with investment financed through borrowing in the same foreign currency. As a result, currency fluctuations will have equal effects on assets and liabilities of the investors, with zero net effect. For instance, from a Japanese investor's viewpoint, a position in Euro Stoxx futures makes or loses money only when the index rises or falls. A change in the yen price of the euro would, by itself, produce neither a gain nor a loss, because the investor has no cash position in euros.

In contrast, the yen return to a fully funded, currency-unhedged investment in Euro Stoxx would be, to a first approximation, the sum of the return on Euro Stoxx, as viewed by a euro-based investor, and the yen return on the euro. Conventional money managers are well aware of the problems raised by currency risk because currency volatility is potentially very large. During periods of increased uncertainty in global markets, the currency volatility may contribute as much to the risk of a fully funded position as the volatility of the underlying asset, the Euro Stoxx.

For CTAs, the only foreign currency risk associated with using futures to trade comes from the value of cash or collateral balances that are the result of either posting margin collateral or accumulating gains or losses in currencies in which the contracts are denominated. Because these balances tend to be small relative to the notional values of the positions taken, foreign currency risk is, for all practical purposes, separate from the risks associated with the underlying assets or commodities. This decoupling allows CTAs to take much more nuanced views on currency exposure than would be possible for most conventional money managers, for whom hedging currency exposure can be costly.

25.5 SYSTEMATIC FUTURES PORTFOLIO CONSTRUCTION

Position taking is a particular feature of futures trading, which differs from investment in traditional assets. In futures markets, one takes positions as opposed to holding underlying assets. The way futures markets approach risk has important implications for the way CTAs do business and for the way someone may choose to invest in these markets. As discussed in the CAIA Level I curriculum, (1) futures markets require gains and losses to be settled in cash daily, (2) futures contracts have no net liquidating value, and (3) futures markets require participants to post collateral to cover potential daily losses. For systematic futures trading portfolios, the key is in building a system for determining positions in futures markets. This section is dedicated to taking a closer look at the basic building blocks in a systematic futures trading system.

25.5.1 The Four Core Decisions of a Futures Trading System

The typical futures trading system is composed of the following four core decisions:

1. Entry: When to enter a position
2. Position sizing: How large a position to take on
3. Exit: When to get out of a position
4. Market allocation: How much risk or capital to allocate to different sectors and markets

Given these four core decisions, a systematic futures trading system is a dynamic system that processes price data inputs, generates trading signals, and outputs automated executable trading decisions. A trading system can simultaneously take into account large amounts of data, process the data, create trading signals, and calculate and allocate risks, as well as determine position sizes, stops, and limits across futures positions. Inside these systems, there are several components that are integrated into portfolio construction: (1) data processing, (2) position sizing, (3) market allocation, and (4) trading execution. Each of these is described in the following sections.

25.5.2 Data Processing in Futures Portfolio Construction

Data inputs for futures trading systems can include both fundamental and technical data. When dealing with futures prices, the aspect of rolling forward futures contracts must be taken into consideration. More specifically, positions will need to be rolled from expiring contracts to newer ones. The rolling aspect of futures contracts creates gaps in price series requiring adjustments around futures expiration dates. Continuous price series are created by removing these gaps.

25.5.3 Position Sizing in Futures Portfolio Construction

Futures trading systems systematically allocate capital to positions across many different asset classes. Position sizing must take into account the volatility of a particular market. One approach to this is **volatility targeting**, where the size of the position is determined by the trader's conviction in her signal, the volatility of the particular futures market, and a volatility target that is determined by the trader. In particular,

$$\text{Number of Futures Contracts} = \text{Sizing Function} \times \frac{\text{Risk Loading} \times \text{Equity}}{\text{Notional Value}} \times \frac{\text{RVol}_T}{\text{RVol}_R} \tag{25.3}$$

where the sizing function reflects the direction of the bet (i.e., long or short) as well as the confidence that the trader has in the signal (i.e., signal strength). For example, a value of 1 indicates that a long position should be taken and that the signal is strong, whereas a value of 0.5 indicates a long position when the signal is not as strong. The risk loading is a parameter selected by traders to reflect the amount of exposure they want to have to the particular market. The value of risk loading is determined by the trader based on the market environment and the amount of risk the trader wishes to take. Notice that risk loading is multiplied by the amount of equity in the portfolio. The risk loading times the equity or capital is sometimes termed the **capital at risk**. For example, if USD 1 million is the available equity or capital, and the risk loading is 0.02, then USD 20,000 is the capital at risk. The denominator is the notional value of the futures contract. The last term on the right-hand side of Equation 25.3 is related to volatility targeting. Here, RVol_T is the realized volatility target, and RVol_R is an estimate of future volatility. This estimate could either be obtained from implied volatility of option prices or be based on realized volatility, calculated using a prespecified window (e.g., 30 daily observations).

APPLICATION 25.5.3a

Consider a CTA with USD 30 million capital. The CTA has determined that 10% of this capital should be allocated to trading in the Brent Crude Oil market. The sizing function is estimated to be 0.8, which means the trader's signal is strong and indicates a long position in this market. The size of each futures contract is 1,000 barrels, and assuming a current price of USD 50 per barrel, the notional value of each contract will be USD 50,000. Finally, assume that the annualized volatility target is 20% and that the annualized realized volatility using near-term futures prices of the past 30 days has been 30%.

Given these figures, the number of futures contracts based on Equation 25.3 will be as follows:

$$\text{Number of Futures Contracts} = 0.8 \times \frac{10\% \times 30{,}000{,}000}{50{,}000} \times \frac{20\%}{30\%} = 32$$

If oil markets become calmer, and the estimate of price volatility declines to 25%, the trader will need to rebalance his position to have 38 contracts in the portfolio.

An alternative approach would be to determine the position size based on a range of factors other than a volatility target. This approach can be expressed as follows:

$$\text{Number of Contracts} = \text{Sizing Function} \times \frac{\text{Risk Loading} \times \text{Capital}}{\text{PVol}_R \times \text{Contract Size}} \tag{25.4}$$

In this expression, the sizing function is similar to what was discussed in Equation 25.3, and it reflects the direction of the bet (i.e., long or short) as well as the confidence that the trader has in the signal (i.e., signal strength). In this case, the risk loading is a parameter selected by the trader to reflect the amount of exposure she wants to have to the particular market and will incorporate a volatility target as well as other information she wants to take into account when determining allocation to this market. Similar to the previous case, the risk loading is multiplied by the amount of equity or capital to determine the allocation to this market. In the denominator of Equation 25.4, PVol_R is daily price volatility of the futures contract, which is multiplied by the size of the futures contract.

APPLICATION 25.5.3b

Continuing with the previous example in which the CTA's capital is assumed to be USD 30 million, suppose the sizing function is 0.8 and the risk loading is 0.2%. The daily price volatility of oil is estimated to be USD 1.1, and each contract is for 1,000 barrels.

Given these figures, the trader will take a long position in futures contracts based on Equation 25.4:

$$\text{Number of Contracts} = 0.8 \times \frac{0.2\% \times 30{,}000{,}000}{1.1 \times 1{,}000} \approx 44$$

In Equation 25.4, the sizing function is multiplied by the total adjusted dollar risk allocated, which is equal to the allocated dollar risk (risk loading times capital) divided by the volatility of price changes measured over the last K trading periods times the point value of the contract. The **point value** is the gain or loss in the contract from a one-point change (e.g., USD 1) in the futures prices. The allocated dollar risk is the amount of capital that is put into active risk, which is the notional amount times the scalar for how much risk will be taken (risk loading). This amount of risk must then be divided by the futures contract dollar risk. The **futures contract dollar risk** is a measure of the riskiness of the underlying asset of the futures contract during the most recent K trading periods and is the denominator in Equation 25.4. It depends on the contract size or point value and volatility of each particular futures contract. It is important to remember that the notional value of one contract is equal to the point value (multiplier) times the contract price. For example, given the size of the Brent Crude Oil futures contract (1,000 barrels) and the price per barrel (USD 50), the notional value of each contract is USD 50,000.

25.5.4 Market Allocation in Futures Portfolio Construction

Market allocation is the process with which both risk and capital are allocated across various futures positions. The process of allocation comes from both capital allocation schemes and risk allocation. Using the equations for the nominal positions in the previous section, there are two avenues by which market allocation can be adjusted based on risk. Risks can be adjusted by the risk loading and the volatility adjustment for risk per contract.

In the simplest case, the risk loading can be set to be equal for all markets. The risk loading is set up in this way to allow for a simple increase or decrease in the overall exposure of the futures trading system. Capital allocation can also vary from market to market. In the simplest case, capital allocation can be equal dollar risk weighted. This means that the capital for an individual market is equal to the total capital divided by the number of traded markets.

Consider a USD 100 million portfolio that is trading 100 futures markets. If each strategy trades equal dollar risk, each market will be allocated 100/100, or USD 1 million. The size of the position for each market will depend on the allocated risk and the amount of realized volatility in each market. This means that the notional exposure in each market can vary substantially depending on each market's volatility. Although the risk allocation will be equal, the notional exposures may differ.

There is a wide range of methodologies for implementing capital allocation. Several of these methods fit easily into the simple structure proposed in this section. Others may require either more complicated or new structures to implement them. The main ways to allocate risk are through equal dollar risk allocation; equal risk contribution, which is similar to risk parity, a topic discussed earlier in this book; and market capacity weighting, in which an allocation is adjusted to reduce the market impact of the futures trading system. In summary:

- **Equal dollar risk allocation** is a strategy that allocates the same amount of dollar risk to each market. This approach does not consider the correlation between markets and is similar to the $1/N$ approach.
- **Equal risk contribution** is a strategy that allocates risk based on the risk contribution of each market, taking correlation into account. This approach is similar to risk parity.
- **Market capacity weighting** is an approach in which capital is allocated as a function of individual market capacity. In futures markets, a market capacity weighting will depend on the market size, as measured by both daily volume and price volatility.

A CTA fund with larger assets under management cannot allocate capital to markets with lower open interest and volumes. These constraints may cause a larger CTA to tend to allocate risk similarly to market capacity weighting.

25.5.5 Trading Execution in Futures Portfolio Construction

The final component of a futures trading system is trading execution. Implementation approaches for turning trading signals into actual positions can vary from one system to another. **Alpha decay** is the speed with which performance degrades as execution is delayed. In the long-term perspective, alpha decay is much less important for trend following than it is for many shorter-term futures strategies. As a result, the more important consideration related to execution for trend-following systems is cost rather than execution speed. Slower futures trading systems create orders that can be executed in a rather passive manner. Some managers may also choose to sample the price throughout the liquid periods of the day to generate signals, and split daily orders into several intraday orders. For the case of trend following, execution is generally done via simple market orders. Stop-loss orders and more complicated limit orders are less commonly used.

25.6 CONCLUSION

Perhaps the toughest question asked by investors who are considering CTAs or managed futures investments for the first time is why CTAs should make money. There are several reasons. If markets are adaptive over time, highly liquid adaptive strategies in futures markets have the potential to find opportunities when markets diverge. CTAs take positions and allocate risk across a wide range of global futures and liquid forward markets. Their dynamic nature allows managed futures strategies to capture divergence and dislocation. Given this, managed futures strategies provide a different and often complementary profile to traditional investment portfolios. Issues related to structuring investments in managed futures will be addressed in the following chapter.

REFERENCES

Aldridge, I. 2009. "Systematic Funds Outperform Discretionary Funds." Working paper, BigDataFinance.org.

Burghardt, G., and B. Walls. 2011. *Managed Futures for Institutional Investors: Analysis and Portfolio Construction*. Hoboken, NJ: Bloomberg Press.

Elton, E., M. Gruber, and J. Rentzler. 1987. "Professionally Managed, Publicly Traded, Commodity Funds." *Journal of Business* 60 (2): 175–99.

———. 1989. "New Public Offerings, Information, and Investor Rationality: The Case of Publicly Offered Commodity Funds." *Journal of Business* 62:1–15.

———. 1990. "The Performance of Publicly Offered Commodity Funds." *Financial Analysts Journal* 46:23–30.

Greyserman, A., and K. Kaminski. 2014. *Trend Following with Managed Futures: The Search for Crisis Alpha*. Wiley Trading Series. Hoboken, NJ: John Wiley & Sons.

Jegadeesh, N., and S. Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48 (1): 65–91.

———. 2011. "Momentum." *Annual Review of Financial Economics* 3:493–509.

Kat, H. 2002. "Managed Futures and Hedge Funds: A Match Made in Heaven." Research paper, Cass Business School, London.

Kazemi, H., and Y. Li. 2009. "Market Timing of CTAs: An Examination of Systematic CTAs vs. Discretionary CTAs." *Journal of Futures Markets* 29 (11): 1067–99.

Lintner, J. 1996. "The Potential Role of Managed Commodity-Financial Futures Accounts (and/or Funds) in Portfolios of Stocks and Bonds." In *The Handbook of Managed Futures: Performance, Evaluation and Analysis*, edited by Carl Peters and Ben Warwick, 99–137. New York: McGraw-Hill Professional. Originally presented to the Analysts Federation in 1983.

Lo, A. 2004. "The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective." *Journal of Portfolio Management* 30:15–29.

———. 2012. "Adaptive Markets and the New World Order." *Financial Analysts Journal* 68:18–29.

Moskowitz, T., T. Ooi, and L. Pedersen. 2012. "Time Series Momentum." *Journal of Financial Economics* 104.

Park, P., O. Tanrikulu, and G. Wang. 2009. "Systematic Global Macro: Performance, Risk and Correlation Characteristics." Working paper. Available at https://www.researchgate.net/profile/Oguz_Tanrikulu2/publications.

Pasquariello, P. 2014. "Financial Market Dislocations." *Review of Financial Studies* 27 (6): 1868–1914.

Schneeweis, T. 2009. "The Benefits of Managed Futures." Benefits Papers Series, Institute for Global Asset and Risk Management.

CHAPTER 26

Investing in CTAs

Managed futures strategies are often systematic strategies that focus on highly liquid and informationally and operationally efficient markets. Futures markets are operationally efficient because of their low transaction costs and bid-ask spreads. This operational efficiency makes these markets informationally efficient as well because a large number of traders with large amounts of capital can participate in these markets. Also, the fact that the underlying assets of futures contracts are major asset classes or economic variables, meaning that a substantial amount of information about them is available to most market participants, contributes to their increased informational efficiency. Trading strategies employed by managed futures traders in these markets lead to a somewhat unique performance profile over time, which has some interesting diversification benefits for both traditional and hedge fund portfolios. This chapter reviews the historical performance of commodity trading advisers (CTAs) and takes a closer look at unique issues for understanding risk in managed futures. Benchmarks, portfolio benefits for including these strategies, and specifics for accessing these strategies are also reviewed.

26.1 HISTORICAL PERFORMANCE OF CTAs

This section provides some evidence regarding the historical performance of CTAs both as stand-alone investment products and in comparison with traditional asset classes (the role of CTAs in diversified portfolios is examined later in this chapter). A closer look at this performance will show encouraging results during periods of market stress. Since systematic trend following is the most popular strategy in the managed futures space, this section also examines this strategy in further detail. For further discussions of historical performances of CTAs and their sources of returns see Elton et al. (1987), Edwards and Park (1996), and McCarthy et al. (1996).

26.1.1 Statistical Properties of CTA Returns

Exhibits 26.1A and 26.1B present basic statistics of two CTA indices—the Barclay Trader Index Discretionary and the Barclay Trader Index Systematic—along with similar statistics for equities, bonds, and commodities. Several issues must be addressed in connection with the results reported here. First, the CTA indices represent equally weighted portfolios of CTAs; therefore, their volatility, skewness, and kurtosis are not necessarily representative of the properties displayed by individual

EXHIBIT 26.1A Estimates of Statistical Properties of Returns on Indices of CTAs and Other Investments

Index (Jan. 2000–Dec. 2014)	Barclay Trader Index Discretionary	Barclay Trader Index Systematic	World Equities	Global Bonds	U.S. High Yield	Commodities
Annualized arithmetic mean	5.8%**	5.5%**	1.7%	6.4%**	7.6%**	7.8%**
Annualized standard deviation	4.1%	8.6%	16.9%	6.2%	11.0%	24.7%
Skewness	0.6**	0.2	−0.6**	−0.1	−1.0**	−0.5**
Kurtosis	1.2**	0.3	1.1**	0.3	6.3**	1.2**
Sharpe ratio	0.8	0.3	−0.1	0.6	0.5	0.2
Annualized geometric mean	5.7%	5.1%	0.3%	6.2%	7.0%	4.8%
Maximum	4.7%	7.4%	11.2%	6.2%	12.1%	19.7%
Minimum	−2.7%	−5.6%	−19.0%	−3.8%	−15.9%	−28.2%
Autocorrelation	−3.1%	−0.2%	18.7%**	5.7%	32.0%**	17.5%**
Maximum drawdown	−4.0%	−10.1%	−54.0%	−10.1%	−33.3%	−67.6%

*Significant at 90% confidence.
**Significant at 95% confidence.
Source: Bloomberg and authors' calculations.

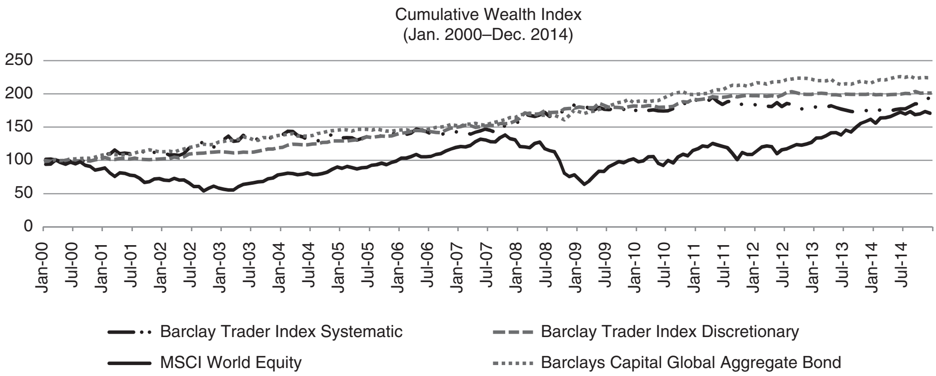


EXHIBIT 26.1B Growth of $100 Invested in CTA Indices and Other Asset Classes
Source: Bloomberg and authors' calculations.

CTAs. For instance, according to Exhibit 26.1A, the annualized volatility of the discretionary index is less than 50% of the volatility of the systematic index. However, if the average volatility of discretionary CTAs is examined, it will be seen that it is only slightly less than the average volatility of individual systematic CTAs. The reason for the difference is this: Because discretionary CTAs are less correlated with each other, a portfolio consisting of a large number of discretionary CTAs will have a relatively small volatility.

Second, both discretionary and systematic CTA indices display zero-to-positive skewness during this period. This is in contrast to the properties displayed by many hedge fund styles and traditional asset classes during the same period. Exhibit 26.2 plots the realized skewness from several hedge fund styles compared with a representative trend-following strategy. Trend-following strategies seem to exhibit positive

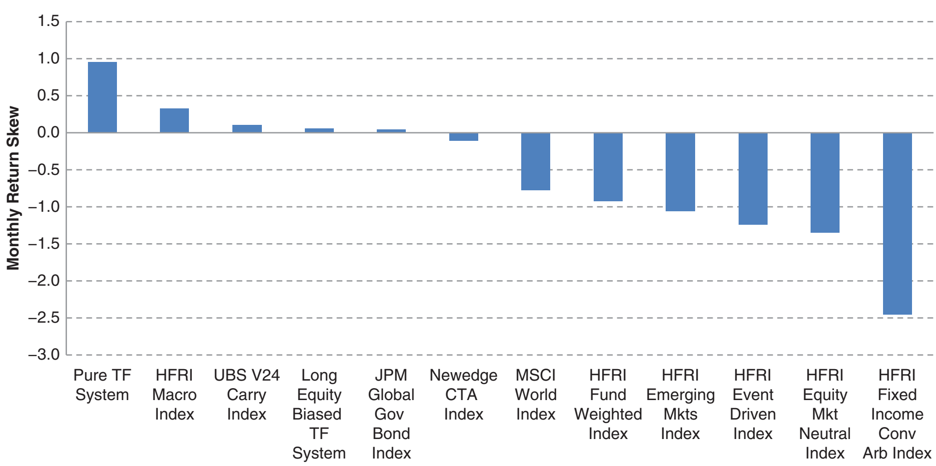


EXHIBIT 26.2 Monthly Return Skewness for Pure Trend Following and Several Other Hedge Fund Strategies and Indices
Source: Greyserman and Kaminski (2014).

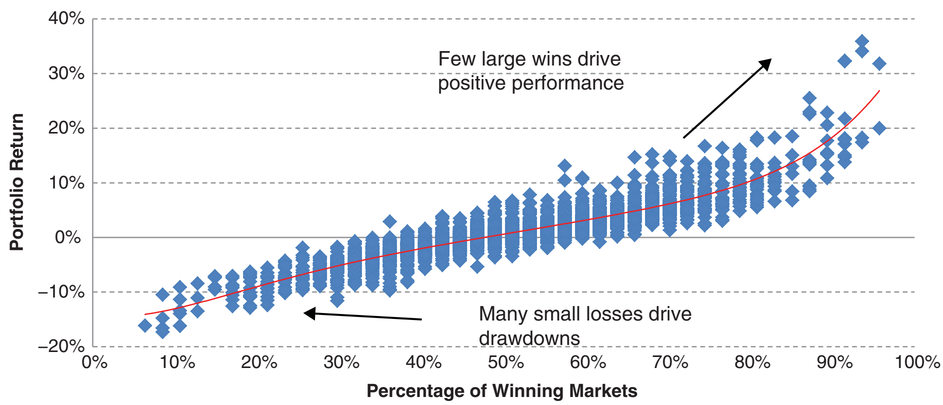


EXHIBIT 26.3 A Scatter Plot between the Portfolio Return and Winning Percentage
Note: This graph uses a 22-day rolling window for a representative pure trend-following system.
Source: Greyserman and Kaminski (2014).

skewness, whereas many other hedge fund strategies maintain negative skewness. Statistical properties of hedge fund returns are examined in Fung and Hsieh (1997) and Kat (2002).

26.1.2 Return Skewness and Trend Following

Diversified trend-following strategies systematically spread risk across many asset classes and cut losses when prices move against their signals. When their signals work, they follow their winners until they begin to revert. As a result of this risk-taking approach, the worst performance comes from many small and additive cutting of losses. Empirically, the **winning ratio of trades**, which is defined as the ratio of the number of profitable trades to the total number of trades, is often less than 50%. In contrast, the best performance comes from finding a trend in prices that tends to result in frequent and large wins. Empirically, both the number of winning trades and the magnitude of the difference between winning trades to losing trades are related to the best performance. Exhibit 26.3 plots the relationship between the percentage of winning trades and trend-following performance. Many small losses across a portfolio cause drawdowns, and somewhat larger, less frequent winners create positive skewness and the convex payout profile seen in Exhibit 26.3. A **convex payout** profile is characterized by an investment strategy that experiences a relatively high frequency of small losses and a relatively low frequency of very large gains. A call option has a convex payout; when the option is in-the-money, it can pay out more than the cost of the option. For trend-following strategies, this means that when there are many small losses, the losses are contained, but when there is a large opportunity, the opportunity pays out more than a linear position would.

26.1.3 Risk Factor Exposure for CTAs

Exhibit 26.4 displays CTA exposure to various risk factors in both multivariate and univariate settings. The important points to take from this exhibit are the low levels

EXHIBIT 26.4 CTA Exposure to Various Risk Factors, January 2000–December 2014

Multivariate Betas	**World Equities**	**Global Bonds**	**U.S. High Yield**	**Commodities**	**Annualized Estimated α**	R^2
Barclay Trader Index Discretionary	0.05*	0.09	−0.12**	0.05**	4.75%**	14.89%
Barclay Trader Index Systematic	−0.02	0.61**	−0.22**	0.09**	2.88%	19.07%
Univariate Betas	**World Equities**	**Global Bonds**	**U.S. High Yield**	**Commodities**	**%Δ Credit Spread**	**%Δ VIX**
Barclay Trader Index Discretionary	0.03	0.08	−0.02	0.05**	0.00	−0.00
Barclay Trader Index Systematic	−0.05	0.53**	−0.14**	0.07**	0.00	0.01
Correlations	**World Equities**	**Global Bonds**	**U.S. High Yield**	**Commodities**	**%Δ Credit Spread**	**%Δ VIX**
Barclay Trader Index Discretionary	0.12*	0.08**	−0.06	0.31**	0.01	−0.08
Barclay Trader Index Systematic	−0.11*	0.28**	−0.17**	0.19**	0.00	0.07

*Significant at 90% confidence.
**Significant at 95% confidence.
Source: Bloomberg and authors' calculations.

of exposure to equity, interest rate, commodity, credit, and volatility risk factors. As will be discussed later in this chapter, these low exposures are the basis of the diversification benefits provided by CTAs. In addition, given the reported R^2 in Exhibit 26.4, we can see that a smaller percentage of volatility of the discretionary index is explained by these factors when compared to that of the systematic index. This confirms that discretionary CTAs are more heterogeneous, and, unlike trend-following CTAs, no single strategy dominates this space.

Finally, Exhibit 26.5 highlights a major benefit of CTAs in providing downside protection during periods of market stress. CTA indices provided positive returns during the post–Internet bubble period (8/2000–12/2002) and at the peak of the financial crisis, represented by the bankruptcy of Lehman Brothers (9/2008–10/2008). For comparison, Exhibit 26.5 plots the performance of CTAs during drawdown periods for equity markets. During equity market drawdowns, CTAs tend to have positive performance.

26.1.4 Evidence on Market Divergence and CTAs

In Chapter 25, market divergence was defined as the process by which market participants and groups of market species evolve and adapt to new market conditions. If markets are adaptive, this will not happen in a frictionless way, and there may be measured trends in financial price data. A period of market stress, such as a financial crisis, provides a clear example of a market disruption, during which many underlying structural forces are changing. Empirically, these periods

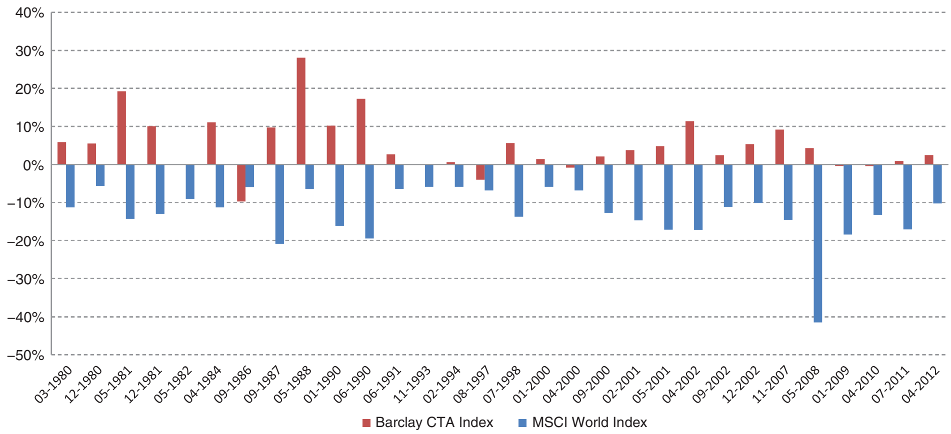


EXHIBIT 26.5 Performance of the Barclay CTA Index during All Drawdown Periods of the MSCI World Index (Dates Reflect Starting Point of All Drawdowns Greater Than 5%)
Source: Kaminski (2011).

can be measured by examining the market divergence index (MDI), as presented in Chapter 25.

Most CTAs follow diversified strategies and attempt to take advantage of opportunities that arise across many futures markets. Therefore, the performance of a typical CTA that pursues a trend-following strategy based on moving average crossovers can be compared with the MDI. Exhibit 26.6 plots market divergence (left-hand side) and trend-following performance (right-hand side). When market divergence is high, this indicates that market prices deviate substantially from a random walk, and there is measurable signal in price series. The trend-following system

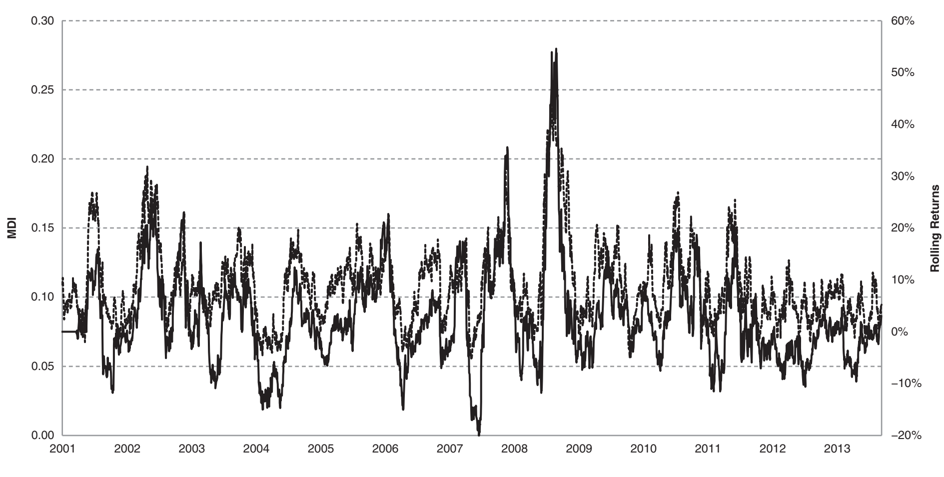


EXHIBIT 26.6 MDI with the Rolling 100-Day Returns of a Representative Pure Trend-Following System, 2001–13
Source: Greyserman and Kaminski (2014).

includes all asset classes: commodities, equity indexes, fixed income, and currencies. In this example, the correlation between market divergence and trend following is 0.74. This high correlation demonstrates the direct link between market divergence and trend-following performance. Performance and divergence tend to be higher during periods of extreme market stress, such as financial crisis. This can be seen by the realized performance of trend following during both the credit crisis and the tech bubble.

CTAs invest over a wide range of highly liquid markets. More importantly, because CTAs do not generally invest in individual securities, market illiquidity and large bid-ask spreads are not significant issues for CTAs during periods of market stress or inefficiency. Conceptually, **market stress** can be defined as a period during which there are larger structural shifts in the financial system, balance in supply and demand, valuation, and aggregate risk appetite. The 2007–9 financial crisis is the most extreme state of market stress.

For the case of market stress, numerous studies have examined the benefits of CTAs in terms of providing downside protection. CTAs have access to multiple highly liquid markets, and they can easily take long or short positions in these markets. Consistent with the adaptive markets hypothesis from Chapter 25, the adaptive nature of these strategies allows them to capture opportunities during periods of market stress. For example, during a financial crisis, a decline in equity prices could benefit CTAs who are (1) short equity indices; (2) long short-term Treasuries, which benefit from a flight to safety; (3) long strong currencies, which typically benefit from turbulence in financial markets; (4) long implied or realized volatility derivatives; and (5) long/short with realignments in currency and commodity markets. For the case of trend-following strategies, the strategy cuts losses but will follow opportunities as they unfold. This means that the strategy will find opportunities as they occur in different asset classes, and each market stress scenario may provide unique opportunities. For the case of trend following, this is consistent with the term *time-series momentum*: the strategy captures momentum across time and across a set of asset classes.

26.1.5 Exposure of CTAs to Market Volatility

This section discusses an important but not well understood question regarding an empirical property of CTAs: Are CTAs long volatility, and, if so, do they benefit from a rise in market volatility? Since some CTA strategies tend to profit from market stress, it is not surprising that CTAs have a complex relationship with volatility. We define a period of market stress as a period during which there is increased volatility and measurable market moves both to the downside for risky investments and to the upside for safe assets. Large market moves that take place as the result of market stress tend to create pockets of market inefficiency and therefore present skilled and well-capitalized investors, including CTAs, with opportunities. During a period of uncertainty without measurable market moves (or inefficiency in prices), volatility will be high, but markets may not be stressed. Increased volatility due only to uncertainty will not provide opportunities for CTAs. In other words, increased volatility that is not accompanied by sustained directional moves in markets does not typically represent a period of market stress. If the volatility of a price series increases but there is no significant directional move (or no trend), a systematic strategy will attempt to follow apparent trends that are not there, leading to moves in and out of long and short positions. This causes losses as opposed to gains.

Given the performance characteristics during crises, some argue that CTAs are long volatility, and since equity volatility appears to be the most dominant factor affecting volatility in other markets, CTAs are expected to benefit whenever equity market volatility spikes. The notion that CTAs are long volatility comes from available empirical evidence that shows that CTAs perform relatively well when the level of equity volatility is high and when equity volatility is increasing (see Exhibit 26.5). Another factor contributing to this issue is a set of academic research showing that CTA returns share similar properties with a long position in a straddle. In a typical straddle, the trader takes long positions in both call and put options on the same underlying asset at the same strike price. The trader will profit if the underlying asset of the straddle makes a large move in either direction. In addition, the sensitivity of the position (i.e., its delta) increases as the underlying asset of the straddle continues its move in the same direction.

The notion that CTAs are long volatility is not very precise. As stated by Malek and Dobrovolsky (2009), volatility exposure would mean different things to traders employing different systematic trading strategies. Furthermore, it is not clear what is meant by "volatility exposure." First, it could mean that CTAs provide higher returns when the level of volatility is high. Second, it could mean that CTAs provide higher returns when the level of volatility is increasing. Equally important, one needs to be precise whether expected or unexpected changes in volatility are being considered. For example, markets tend to become more volatile around major monetary policy announcements by a central bank. Such an increase in volatility is expected and is therefore not likely to present trading opportunities for CTA strategies. Therefore, the way that volatility is measured has important implications for estimating the volatility exposures of CTAs.

26.1.6 The Gamma Exposure of CTAs

To gain a better understanding of the volatility exposure of CTAs, terminology from option pricing is used to analyze the behavior of a typical trend-following CTA. Suppose the market is directionless and a CTA has no positions. In this case, the CTA's directional market exposure (i.e., its delta) is zero. Once the market starts to move and begins a trend, the CTA begins adding positions. When the market moves up, the CTA increases its long position (i.e., its delta becomes positive and increases), and when the market goes down, the CTA increases its short position (i.e., its delta becomes negative and decreases). This behavior is exactly what makes the return profiles of trend-following CTAs similar to those of a long position in a straddle. However, this return profile is not the result of any exposure to volatility; rather, it is the result of being long gamma. Using option terminology, gamma measures the rate of change in the delta of an option as the price of the option's underlying asset changes. This means that the delta of a position that is long gamma increases as the price of the underlying asset increases, and decreases as the price of the underlying asset decreases. Therefore, trend-following strategies are characterized by being long gamma, and relative value strategies are characterized by being short gamma. As a result, even though return profiles of trend-following CTAs are similar to return profiles of straddles, it does not follow that CTAs are long volatility simply because straddles are long volatility. One could argue that CTAs have the return profile of being long calls and put options, but this profile is not created by taking long positions in volatility; rather, through their dynamic trading, CTAs attempt to create a return profile that is similar to a long gamma strategy.

Option premiums are most sensitive to volatility when the options are at-the-money. When they are deep-in-the-money, they behave similarly to long or short positions in their underlying assets; when they are deep-out-of-the-money, they become almost worthless. Thus, a straddle displays volatility exposure only when it is close to being at-the-money. As the underlying price moves away from its strike level, the straddle's sensitivity to volatility declines. This means that a trend-following CTA that is trying to take advantage of a trend behaves like an at-the-money straddle and, as a result, can be characterized as being long gamma. We define a **dynamic-trading-based long gamma strategy** as a portfolio management method that modifies portfolio weights through time using a method that causes relatively high probabilities of relatively small losses and relatively low probabilities of relatively large gains, resulting in a convex payoff profile in Exhibit 26.7.

Exhibit 26.7 displays the relationship between the CISDM CTA Equal Weighted Index return and its volatility during specific six-month periods. For example, from the top part of the exhibit we can see that in the six-month period leading up to February 2009, the MSCI World Index lost an average of 8.8% per month. During the same period, the CISDM CTA index earned 1.6% per month. The bottom part of the exhibit displays similar information, but it is sorted based on the average volatility of the MSCI World Index. For example, during the sixth-month period leading up to June 2004, the MSCI World Index had a very low monthly volatility of 1.6%. During the same period, CISDM CTA index had an average volatility of 2.5%, while losing 0.3% per month. We can see that during periods when equity volatility is high, the CTA index makes money while equity markets lose money. However, the opposite is not necessarily true. That is, the CTA index does not consistently lose money when the MSCI World Index's volatility is low. This provides evidence supporting our argument that CTAs are not long volatility but rather long gamma, and are therefore dynamic-trading-based long gamma strategies.

26.2 DIVERSIFICATION BENEFITS OF CTAs

Diversification, the act of introducing imperfectly correlated returns into an investment portfolio, is often touted as the best method for achieving some amount of protection during periods of distress. This section examines a few desirable diversification measures for portfolios, and demonstrates how adding CTAs using a representative trend-following strategy can be a proxy to a 60/40 equity/bond and a fund-of-funds (FoF) portfolio.

26.2.1 Crisis Alpha and CTA Performance

Crisis alpha is the measure of a strategy's performance during market stress. Crisis alpha is one of the most important portfolio benefits of trend-following strategies. Exhibit 26.8 shows the crisis alpha for pure trend following and several other hedge

EXHIBIT 26.7 Historical Performance of CTAs and Global Equity Markets and Their Corresponding Volatilities

	Worst 6-Month Period in Terms of MSCI Return						Best 6-Month Period in Terms of MSCI Return					
Monthly Figures	**Feb-09**	**Nov-08**	**Jan-09**	**Oct-08**	**Dec-08**	**Mar-09**	**Dec-10**	**Oct-09**	**Feb-11**	**Jul-09**	**Sep-09**	**Aug-09**
MSCI World return	–8.8%	–8.2%	–7.3%	–6.8%	–6.3%	–5.6%	3.8%	4.0%	4.0%	4.3%	6.1%	6.7%
CISDM CTA return	1.6%	1.1%	1.5%	1.2%	0.9%	1.2%	2.3%	0.3%	2.0%	–0.2%	0.4%	–0.1%
MSCI World volatility	11.0%	10.1%	10.2%	9.8%	9.6%	10.4%	6.3%	5.7%	5.5%	8.6%	7.3%	7.7%
CISDM CTA volatility	2.6%	3.2%	2.6%	3.3%	3.0%	2.7%	3.7%	1.9%	3.5%	1.8%	1.8%	1.9%
	Highest 6-Month Period in Terms of MSCI Volatility						**Lowest 6-Month Period in Terms of MSCI Return**					
Monthly Figures	**Feb-09**	**Mar-09**	**Jan-09**	**Nov-08**	**Oct-08**	**Dec-08**	**Jan-15**	**Dec-12**	**Dec-14**	**Aug-04**	**Aug-14**	**Jun-04**
MSCI World return	–8.8%	–5.6%	–7.3%	–8.2%	–6.8%	–6.3%	–0.2%	1.6%	–0.1%	–0.4%	1.0%	0.6%
CISDM CTA return	1.6%	1.2%	1.5%	1.1%	1.2%	0.9%	3.0%	–0.2%	2.2%	–1.4%	0.7%	–0.3%
MSCI World volatility	11.0%	10.4%	10.2%	10.1%	9.8%	9.6%	1.9%	1.9%	1.9%	1.8%	1.6%	1.6%
CISDM CTA volatility	2.6%	2.7%	2.6%	3.2%	3.3%	3.0%	3.5%	1.8%	2.9%	1.9%	1.6%	2.5%

Source: Bloomberg and authors' calculations.

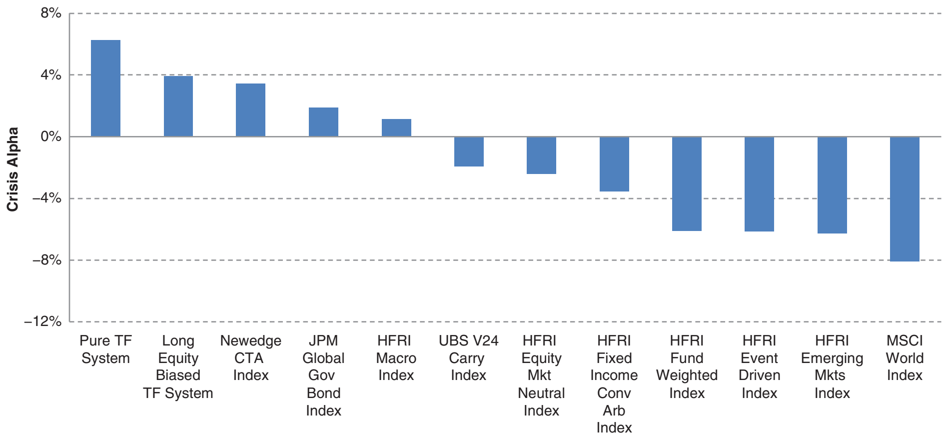


EXHIBIT 26.8 Crisis Alpha for Pure Trend Following and Several Other Hedge Fund Strategies and Indices
Source: Greyserman and Kaminski (2014).

fund strategies and indices during equity market crises.[1] Although most other hedge fund strategies and indices provide negative crisis alpha, the pure trend-following program provides a monthly crisis alpha of 6%. Given the performance of most traditional portfolios during a market crisis, crisis alpha provides substantial diversification benefits for an institutional investor.

A number of sources contribute to the crisis alpha for CTAs. First, CTAs trade in the most liquid markets; therefore, their trading strategies are not negatively affected by the lack of liquidity that seems to characterize periods of market stress. Second, unlike many hedge fund and mutual fund strategies, which tend to focus on only one asset class, CTAs can freely allocate to various asset classes, which allows them to take long positions in asset classes that are most likely to experience a gain as a result of increased financial distress, or short positions if expecting a loss. Third, strategies that trade individual securities may find it difficult to short certain securities and could be subject to a short squeeze. On the other hand, CTAs trade financial futures, which are not subject to short sale restrictions and short squeezes.

26.2.2 CTA Diversification Benefits for a 60/40 Investor

A 60/40 portfolio is any portfolio that can be reasonably well described as having risk exposures approximating 60% in equities and 40% in bonds. The 60/40 investor is represented by combining a 60% allocation to the MSCI World Index and a 40% allocation to the JPMorgan Global Bond Index. Exhibit 26.9 depicts the five-year rolling Sharpe ratio of the 60/40 portfolio combined with a 20% allocation to a representative trend-following system.

We can see that while CTAs may not represent an attractive investment on a stand-alone basis, they provide a meaningful amount of diversification benefits when combined with a 60/40 portfolio.

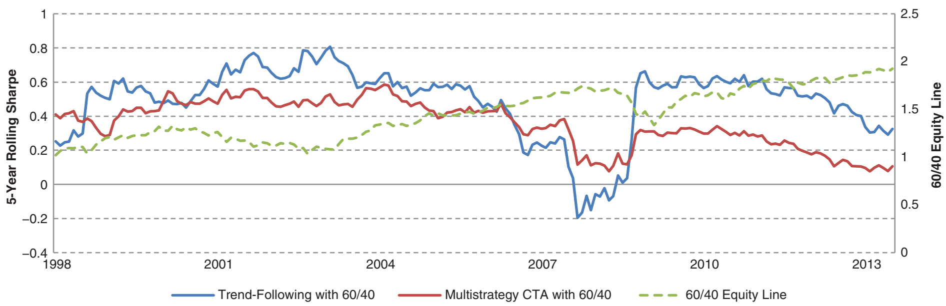


EXHIBIT 26.9 Five-Year Rolling Sharpe Ratio for a Representative Trend-Following Strategy Plotted Together with the Cumulative Performance of Traditional 60/40 Portfolios, 1998–2013
Source: Greyserman and Kaminski (2014).

26.2.3 CTA Diversification Benefits for a Fund-of-Funds Investor

Using the HFRI FoF index as a proxy for the portfolio of a fund-of-funds investor, Exhibit 26.10 displays the five-year rolling Sharpe ratio of a portfolio combining a traditional fund-of-funds investment with an allocation to a CTA in an 80/20 ratio.

Similar to the information provided by the previous exhibit, we can see that CTAs provide certain benefits to FoF investors. In particular, we clearly see that the Sharpe ratio of the CTA index and the FoF index move in opposite directions and this relationship becomes stronger during periods of financial distress. While CTAs may follow strategies that overlap significantly with some hedge fund strategies, the fact that they implement these using highly liquid markets allows them to implement their strategies even when market liquidity is low in traditional cash markets.

26.3 CTA RISK MEASUREMENT AND RISK MANAGEMENT

CTAs invest capital by allocating risk across futures markets using primarily futures contracts and highly liquid interbank forward contracts. Many investment strategies

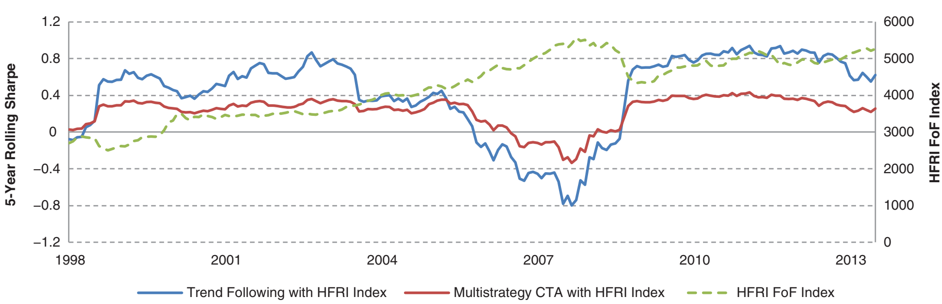


EXHIBIT 26.10 Five-Year Rolling Sharpe Ratio for a Fund-of-Funds Portfolio Combined with 20% of the Representative Trend-Following Strategy Plotted with the Cumulative Performance of the HFRI FoF Index, 1998–2013
Source: Greyserman and Kaminski (2014).

invest all of their capital directly into different assets. When an asset is purchased, the asset is delivered into the portfolio in exchange for cash. In this case, the owner of the portfolio takes ownership of the assets in the portfolio. Managed futures strategies rely exclusively on the use of derivatives contracts. As opposed to allocating notional capital to hold assets in different asset classes, they allocate a large amount of notional capital to low-risk assets (such as Treasury bills) and use the remaining notional capital to fund active risk exposures in a wide array of asset classes. In this sense, one can think of CTAs as having a risk budget (e.g., volatility of the portfolio), which they have to decide how to allocate among various asset classes. Thus, for a CTA, allocating the risk budget is unrelated to how capital is allocated.

Risk is allocated by taking positions (or exposures) in different futures markets. These exposures are funded by margin (as collateral) with a central clearinghouse. Note that whereas traditional investment portfolios are all maintained in brokerage accounts, managed futures portfolios have both low-risk funding portfolios and active-risk trading portfolios.

26.3.1 Leverage versus Implicit Leverage

Leverage allows a portfolio to magnify its return and risk exposure. In traditional investment strategies, leverage is achieved by borrowing capital or obtaining credit from a broker or a bank. This capital is then used to purchase assets—or, in the case of a short sale, assets are lent out in exchange for cash. These transactions occur between two counterparties: the investor and the broker (for example). There are many concerns with leverage using traditional assets. First, leverage is costly. Since traditional portfolios require taking delivery of assets, borrowing restricts the use of capital and requires interest payments to compensate for the loss of use (short rebate or appropriate loan rates). Second, since all assets are used to collateralize leveraged portfolios, the losses can be magnified and directly cause losses to the entire portfolio. Third, there are asymmetric constraints on leverage between short and long. Shorting is more difficult and costly. For these reasons, in the United States, Regulation T limits leverage for long positions at 50% notional and 150% notional for short positions. Fourth, since leverage requires the borrowing or lending of capital or securities, the contract between the investor and the broker has counterparty risk.

For futures portfolios, the vast majority of the notional capital is invested in low-yield government securities or even cash, and is held in a funding account. This low-risk investment is coupled with active risk investments in futures contracts. Each futures contract gives an investor implicit leverage, which is different from leverage in traditional investment portfolios.[2] **Implicit leverage** equals the notional value of the position divided by the initial margin. This implies that the maximum potential profit or loss exposure of a position is greater than the capital at risk. For example, suppose a fund has 5% margin on long oil futures contracts for 100,000 euros. The notional capital exposure in the futures contract is 100,000 euros, but the notional capital at risk for this position is only 5,000 euros. This means the implicit leverage of this position is 20.

It is important to note that CTAs do not employ leverage in the traditional sense of borrowing money to increase exposure. Capital that is not being used as margin can be invested in Treasury securities or various liquid instruments that meet the requirements of the clearinghouse and the customer's broker. When leverage is applied in traditional portfolios, the assets are used as collateral for the portfolio, and all notional capital is exposed to spiraling losses. This is precisely why managed futures portfolios often impose risk limits per market and cap the amount of notional capital they will continue to expose to futures market risks.

The futures exchange and its clearinghouse provide a process through which buyers can meet the sellers while avoiding the counterparty risk associated with such transactions. In the absence of futures exchange, it is still possible to have a forward contract between two parties, but the credit risk of the counterparty is a major issue. The futures exchange and its clearinghouse record every transaction, facilitate and regulate the delivery of goods, and settle all trading transactions. The clearinghouse plays a third-party role in every transaction that takes place. It assumes a role between the buyer and the seller, with the economic effect being that the seller has sold the contract to the clearinghouse and the buyer is purchasing the contract from the clearinghouse. The net effect is that the numbers of contracts bought and sold are the same. The clearinghouse is able to ensure that both parties meet their dues through the use of margin payments. The traders make margin payments to the clearinghouse, which are then used to ensure that buyers and sellers are able to meet their obligations.

Futures contracts provide exposure to moves in the prices of a wide range of underlying assets. Futures markets are built by bringing together pools of investors both long and short who pool their collateral to maintain their views of each asset. For example, investors who hold a long position in the S&P 500 futures contract will profit if the S&P 500 goes up in value. Their gain will be at the expense of the investors who hold a short position in the S&P 500 futures contract. The clearinghouse uses the margin accounts to make offsetting entries in buyers and sellers of futures contracts each day. This means all gains and losses are settled through the clearinghouse at the end of the day.

26.3.2 Margin Accounts and Collateral Management

Futures markets require participants to post collateral to cover potential daily losses. The posting of collateral is typically called margin. The minimum margin for a globally diversified CTA portfolio would most likely be enough to meet this objective, since margin is only aggregated within each clearinghouse. Globally diversified CTAs trade across many clearinghouses and do not often get cross-margin benefits across exchanges. A cross-margin benefit is available when a CTA has multiple positions in contracts traded on the same exchange, which allows the total amount of margin that the CTA must post to be smaller than the sum of the margin amount of each contract. In practice, each clearinghouse requires that margins be posted in the local currency, causing CTA portfolios to be exposed to currency risks. To alleviate this problem, many CTAs can use single-currency margining, in which the trading client can post the full margin in the form of dollars (or euros or any other allowed currency). Under this arrangement, the clearinghouse is responsible for converting the client's cash into collateral that is acceptable for various exchanges around the world.

As previously discussed, for a well-diversified CTA portfolio, the passing of each trading day produces gains and losses in various currencies. In futures markets, it is standard practice to settle all gains and losses in cash every day. These daily cash settlements of gains and losses are known as variation margin. The practice of making daily cash settlements produces an ongoing stream of small transaction costs. Minimizing these costs is an important objective for CTAs and their investors. The cash that flows into or out of an account also affects the ultimate return earned by the investor. Cash that flows in can be invested, whereas cash that flows out must be financed, either explicitly or out of pocket.

Because all gains and losses are settled in cash daily, futures contracts have no net liquidating value beyond what they accumulate over the course of a single trading day. As a result, there is no natural denominator for estimating the return on a futures position. Interest earned on any cash or collateral invested in a fund or posted as collateral in a managed account is part of the total return. The investor must always care about the security of the cash investment.

To cope with these specialized features of futures markets, the CTA industry has adopted three specialized terms that investors must know: (1) trading level, (2) funding level, and (3) notional level.

Trading level is simply the base or denominator used in calculating returns of leveraged positions. It is the mutually agreed upon amount to be traded that determines the size of the actual positions that the CTA takes in futures markets, depending on the CTA's leverage goals. It is the account value that the CTA uses to translate futures profits and losses into percentage returns. It is also the account value that the CTA uses to calculate management and incentive fees. The trading level is the amount of capital that is traded in the active risk account.

The **notional funding** gives investors the ability to leverage their managed futures account to a higher trading level than would exist with cash funding. Notional funding in managed futures is favored by investors because it capitalizes on the ease of acquiring leverage through futures markets. In addition, implicit leverage may carry a relatively low cost because the notionally funded amount is not borrowed or deposited; the funding level is a good-faith deposit for the full value of the account.

For example, if an investor wanted to invest with a CTA that required a minimum investment of \$500,000, the investor could either fully fund the account with \$500,000 or, if notional funding was offered, partially fund the account (e.g., \$250,000) but still have it traded as if it were funded with \$500,000. In this case, the trading level—which is also the amount on which investor fees and returns are calculated—would be \$500,000 with the account funded 50%. If the CTA returned 10% that year, the investor would have made \$50,000 (a 10% gain on the trading level), but it would be a 20% gain on the notional funding level. Of course, if the CTA were to lose 10%, the loss would be magnified to 20%.

As the preceding example demonstrates, the **funding level** is the total amount of cash or collateral that the investor posts or invests to support the trading level. The rock-bottom minimum funding level for any futures position or portfolio is the total value of margin collateral required by the various futures exchanges. The relationship is depicted in Equation 26.1:

$$\text{Trading Level} = \text{Funding Level} + \text{Notional Level} \tag{26.1}$$

Exchange margins tend to be small relative to the face value or portfolio equivalent value of the underlying contracts. In a diversified portfolio of futures contracts, the actual day-to-day risk in the portfolio can be smaller still. As a result, funding levels can be lower than trading levels. If they are, the difference between the trading and funding levels is known as the **notional level**, or the amount being notionally funded.

APPLICATION 26.3.2

Consider an account with $100,000 cash invested but with a mandate that risk exposures, fees, and returns be based upon $200,000 through the use of leverage. Identify the trading level, funding level, and notional level.

Equation 26.1 provides the relationship between the three levels. The $100,000 is the funding level, since it is the cash contributed. The $200,000 meets the definition of the trading level. Therefore, the $100,000 being used to lever the funding level to the trading level represents the notional level.

Because futures contracts are exchange-traded instruments, CTAs need to satisfy the rules and restrictions put forth by these exchanges. One of these rules pertains to the amount of collateral or margin that a party to a futures contract must put up. The amount of initial margin is the amount of cash or Treasury bills that must be in an account at a broker or futures commission merchant in order to initiate a trade in a specific futures contract. Minimum initial margins are set by the exchanges for each futures contract. This initial margin, which is generally only a small percentage of the notional value of the futures contract, is related to the volatility of the assets underlying the futures contract and can change over time. More volatile contracts require larger margins. Futures exchanges have the ability to change margin requirements at any time. Margin requirements are often increased after a sharp rise in prices or price volatility. The maintenance margin, which is typically lower than the initial margin, is the amount of margin required to carry previously initiated positions. If a customer's margin account drops below the level required for the maintenance margin, then the customer has to add funds to restore the margin account to the level of the initial margin.

In certain circumstances, the initial margin may be less than the sum of the initial margins of the individual futures contracts. Futures exchanges take into account the fact that the manager is holding both long and short positions in related contracts. Since such spread positions may have less risk than outright directional positions, the exchanges apply a lower spread margin. For instance, a relative value trader might be long a nearby futures contract (e.g., long March corn) and short a distant contract (e.g., short September corn). In such cases, spread margins would apply.

The **margin-to-equity** ratio is expressed as the amount of assets held for meeting margin requirements as a percentage of the net asset value (NAV) of the investment account. For example, if the equity invested in a futures portfolio is valued at $1,000,000 and the total margin required by various exchanges is $61,000, then the margin-to-equity ratio would be 6.1%. It is often difficult to interpret margin-to-equity ratios. High levels could indicate highly levered trading; for instance, a margin-to-equity ratio of 100% means that the invested equity is just enough to cover the margins, indicating that the maximum amount of leverage provided by the contracts is being employed. However, high levels of the margin-to-equity ratio could equally result from a portfolio diversified across many futures markets where margin offsets are not available (e.g., partially offsetting positions at different exchanges). Still, the margin-to-equity ratio is a popular measure of CTA risk because it tells investors roughly how much of their investment could be used for margin purposes. This number will fluctuate from day to day for a given manager, but investors can obtain the average range. It should be noted that the exchange sets the margin to reflect the riskiness of the futures contract. The greater the exchange's estimate of the risk associated with a contract, the higher the level of the margin required by that exchange on the particular contract.

26.3.3 Capital at Risk for Managed Futures

Managed futures traders normally employ stop-loss rules in their trading programs. These stop losses are specific prices at which the strategy will exit a futures position should the price move adversely. In the case of reversal systems, stops are effectively the price at which the system liquidates an existing directional position and establishes a new position in the opposite direction. Capital at risk (CaR) represents the total loss that would be incurred should each position hit its stop-loss price level on that day. Exhibit 26.11 displays the CaR of a sample portfolio of long futures positions. It assumes that each stop loss is set at 1% of the notional value of each contract (i.e., the position would be liquidated upon a 1% adverse price move).

The usefulness of CaR is dependent on the stop-loss level established for individual futures positions. If this level is very close to the current market price, the CaR might underestimate the real risk of loss, since unanticipated price volatility could lead the futures price to gap through the stop-loss level, resulting in a greater loss than that being reported by the CaR. This is a particularly important risk in less liquid markets, such as futures on agricultural commodities. In another sense, though, CaR often overstates a portfolio's risk, since it does not account for the possibility that a portfolio might hold both long and short futures positions, which could offer some

EXHIBIT 26.11 Leverage and Capital at Risk (CaR) of a Sample Portfolio

Contract	Notional Contract Value	Loss at 1% Price Change
S&P 500 Index	$ 517,250	–$ 5,172
Corn	$ 19,060	–$ 191
Soybeans	$ 49,500	–$ 495
Eurodollars	$ 987,000	–$ 9,870
U.S. long bond	$ 155,000	–$ 1,550
Crude oil	$ 48,800	–$ 488
Gold	$ 108,000	–$ 1,080
Japanese yen	$ 154,800	–$ 1,548
Size of notional positions	$2,039,410	
Total CaR value		**–$20,394**
Assumed account value	$1,000,000	
Notional leverage	203.94%	
Capital at risk	–2.0394%	

Source: Authors' calculations.

offset should significant price moves occur. That is, it is unlikely for all positions to hit their stop-loss levels simultaneously. Finally, a typical investor does not have the transparency to the CTA's positions that is needed to calculate CaR. The exception is when the investor uses a managed account platform to invest in a CTA program. Under the managed account platform, the investor has access to the positions and can therefore calculate the CaR of a portfolio. The advantages and disadvantages of managed account platforms and other approaches to accessing CTA programs are discussed later in this chapter.

26.3.4 Value at Risk for Managed Futures

Value at risk (VaR) is a method of measuring the potential loss in an investment portfolio given a particular holding period, with no changes to the portfolio during the holding period, and at a particular confidence level. The most common confidence levels used are 95% and 99%. A portfolio's one-day VaR of $3 million at a 95% confidence level means that there is a 95% probability that losses sustained by the portfolio over the next day will not exceed $3 million and, thus, a 5% chance that losses will be greater than $3 million. In other words, during 100 trading days, losses exceeding $3 million are expected in only five trading days. If the value of the portfolio is $100 million, then the VaR could be expressed as 3%.

For example, using the portfolio shown in Exhibit 26.11 (i.e., one long contract in each of the individual markets), a one-day VaR can be computed over a continuous 60-day period. At a 99% confidence level, the one-day VaR could be $9,453, or approximately 1% of the assumed account value of $1,000,000. That is, one would expect that 99% of the time, the daily loss on this portfolio would be less than $9,453.

There are several methods for estimating the VaR of an investment. The simplest and most common method is the parametric approach. In the parametric approach, one assumes that return on the investment follows a known distribution (typically normal). In order to estimate the VaR of the investment, one has to estimate the parameters of the distribution (typically, mean and standard deviation). In fact, to obtain an accurate estimate of VaR, it is crucial to obtain an accurate estimate of return volatility. There are several methods for estimating return volatility, and, in general, the higher the frequency of available observations, the more accurate the estimated volatility.

The most common method for estimating volatility is to obtain daily returns on an investment, say a CTA, and then perform the following procedure on the data to obtain an estimate of daily volatility:

$$\mu = \frac{1}{T}\sum_{t=1}^{T} R_t \tag{26.2}$$

$$\sigma_T^2 = \frac{1}{T-1}\sum_{t=1}^{T} (\mu - R_t)^2 \tag{26.3}$$

Here, σ_T^2 is the estimate of the current daily variance, R_t is the daily rate of return, T is the number of observations, and μ is the estimated daily mean return. In this method, all observations have the same weight. An alternative to this method is to assign larger weights to the most recent observations and smaller weights to observations that occurred many periods ago. One simple and popular technique that uses this alternative weighting is the exponential smoothing method. Under this approach, the estimated value of the variance is obtained using the following expression:

$$\mu_{T-1} = (1 - \lambda)\mu_{T-2} + \lambda(\mu_{T-2} - R_{T-1}) \tag{26.4}$$

$$\sigma_T^2 = (1 - \lambda) \times \sigma_{T-1}^2 + \lambda(\mu_{T-1} - R_T)^2 \tag{26.5}$$

Here, $0 < \lambda < 1$ is a parameter that is selected by the user. Notice that the mean daily return is updated as new observations of returns are received. The updated mean is then used to calculate the daily variance. The larger the value of λ, the higher the weight assigned to the most recent observations in calculating both the mean and the variance.

Once an estimate of the standard deviation is obtained, the user needs to specify the desired level of confidence. The higher the selected level of confidence, the higher the estimated value of VaR will be. After the desired confidence level is selected, the critical value of α is calculated as follows:

$$\Pr\{Z \le \alpha\} = 1 - \text{Confidence Level} \tag{26.6}$$

Here, Z is a standard normal random variable. This means α is the value for a standard normal random variable where the probability of observing a value less than α is equal to 1 minus the selected confidence level. To determine the value of α, one may use either a table of standard normal random variables or a spreadsheet program. Once this critical value is calculated, the VaR of the portfolio can be estimated using the following expression:

$$\text{VaR}_\alpha = \alpha \times \sigma_t + \mu \tag{26.7}$$

The VaR obtained using this expression is typically a negative number. However, it is customary to report the absolute value of this figure. In addition, since the daily mean, μ, is likely to be very small, it is common practice to ignore it.

For example, assume that the daily returns for a CTA are used to obtain estimates of daily volatility, σ_t, and daily mean return, μ, of the CTA. These are reported to be 1.8% and 0.03%, respectively. What is the daily VaR of this CTA at the 95% confidence level?

First, the critical value of α for the 95% confidence level is −1.6448. That is, $\Pr\{Z \le -1.6448\} = 1 - 0.95$. Thus, the daily VaR at the 95% confidence level will be calculated as follows:

$$\text{VaR} = -1.6448 \times 1.8\% + 0.03\% = -2.93\%$$

There is 5% probability that the daily loss experienced by this CTA will exceed 2.93%.

One of the critical assumptions of using this approach to calculate VaR is that the probability distribution of the portfolio's daily return is approximately normal.

If the probability distribution deviates from normality, then alternative methods to estimate the VaR must be used.

As a method of calculating risk, VaR is useful but should be used in conjunction with additional risk-measurement techniques. Its reliance on specific estimates of correlations and volatilities makes it prone to underestimating potential tail risk during periods of increased financial stress.

26.3.5 Drawdown and Managed Futures

Maximum drawdown is calculated as the relative value of the last peak NAV to the all-time low NAV since the peak was reached. For example, using performance figures covering January 2008 through December 2014, the maximum drawdown for a CTA can be estimated using the following expression:

$$\text{Maximum Drawdown} = \left[\min\left(\frac{NAV_{t+i}}{NAV_t}\right) - 1\right] \times 100 \qquad (26.8)$$

where t is any date between January 2008 and December 2014, and $t + i$ is any date between t and December 2014. For example, suppose the NAV of a CTA at the end of June 2009 (time t) was 120, and the NAV of the same fund was 90 at the end of August 2010 (time $t + i$). The ratio would be (90/120) = 0.75. Suppose this figure turns out to be the lowest value that can be obtained using the fund's NAVs between January 2008 and December 2014. In this case, the maximum drawdown of the fund is calculated to be –25%.

Maximum drawdown is a useful measure of risk because it shows the extent of the losses sustained by a manager, and may be used to determine if the manager has the skills to reevaluate the model and make appropriate adjustments when losses are increasing. Risk managers may also use maximum drawdown. For example, a risk manager may follow the maximum drawdown of a fund and decide that whenever the fund's maximum drawdown reaches some predetermined level, the fund should significantly reduce its leverage or reevaluate the fund's trading system.

Many factors affect the maximum drawdown of a manager. The most important factors are volatility of the returns, mean of the returns, and length of time during which the maximum drawdown is calculated. In particular, everything else being constant, maximum drawdown will be higher for CTAs with higher return volatility, whereas maximum drawdown will decline as the average return on the CTA increases. Holding everything else constant, there is greater chance of a large drawdown as longer time periods are used to calculate drawdowns.

Related to maximum drawdown is the **drawdown duration**, which refers to the period between two peaks in the NAV (i.e., the amount of time it takes to go from one NAV high to another NAV high). A short drawdown duration means that the CTA tends to recover quickly from losses. To gain further insight, one can calculate the **maximum drawdown duration**, which measures the longest value of all the drawdown durations during a given period. This figure indicates the longest period of time it took the CTA manager to recover from losses. A large maximum drawdown is a sign that the manager's risk controls were not effective in preventing large losses and that they were slow to react to changes in market conditions. In addition, whereas investors want CTAs to reduce their risk allocation when losses begin to accumulate, they want CTAs to ramp up their allocations once market conditions become more favorable. A CTA manager who can recognize changes in market conditions quickly should have a relatively small maximum drawdown and a short maximum drawdown duration.

26.3.6 Simulation Analysis and Managed Futures

A **stress test** or **scenario analysis** is a market simulation applied to a portfolio to determine how it will perform under different market scenarios. Commonly, these try to focus on extreme market events, both those historically encountered (e.g., the financial crisis in the fall of 2008) and those based on simulated financial stress. This technique is often used in conjunction with VaR, since it examines scenarios in which volatility and correlations are assumed to change.

Exhibit 26.12 illustrates a stress test under simplified assumptions. It analyzes the demand on cash in an investment account should a set of futures positions have an eight-standard-deviation adverse price move and, at the same time, should futures exchanges double the required initial margin on these same positions. For example, previous calculations may have shown that this portfolio has a margin-to-equity

EXHIBIT 26.12 Portfolio Stress Test: Eight-Standard-Deviation Price Move; Doubling of Initial Margin

Contract	Notional Contract Value	1-Standard-Deviation Price Move	8-Standard-Deviation Price Move	Initial Margin Requirement
S&P 500 Index	$ 207,250	1.24%	−$20,559	$30,938
Corn	$ 17,913	1.68%	−$ 2,408	$ 2,025
Soybeans	$ 47,475	1.45%	−$ 5,507	$ 4,725
Eurodollars	$ 987,650	0.08%	−$ 6,321	$ 1,485
U.S. long bond	$ 126,640	0.63%	−$ 6,383	$ 4,320
Crude oil	$ 40,320	2.17%	−$ 7,000	$ 8,100
Gold	$ 90,166	1.26%	−$ 9,089	$ 5,399
Japanese yen	$ 139,636	0.66%	−$ 7,373	$ 4,860
Total loss from 8-standard-deviation price move			−$64,640	
Total required initial margin				$61,852
Potential Total Cash Demand				
8-standard-deviation price move	$ 64,640			
Initial margin at 2× level	$ 123,704			
Total	$ 188,344			
Assumed account value:	$1,000,000			
Potential total cash demand as % of account value	18.83%			

Source: Authors' calculations.

ratio of 6.1%, a CaR of 1.66%, and a VaR of 1%. However, under the conditions assumed in the stress test, the portfolio could lose over 6.46% on a day with an eight-standard-deviation price move (6.46% = 64,640/1,000,000). This is significantly higher than indicated by the CaR or the VaR. In addition, the doubling of the initial margin creates a further demand on cash of 12.4% of the account value. The combined impact would be an 18.8% use of cash in the portfolio over one day. Although this may be an unlikely scenario, the purpose of stress tests is to examine the potential impact of low-probability events. It has been seen in the past that these low-probability events do occur (e.g., failures of Long-Term Capital Management and Lehman Brothers). Similar to the CaR calculation, an investor who has allocated to more than one CTA is not in a position to calculate these stress figures unless the investments are made using managed account platforms.

26.3.7 The Omega Ratio and Managed Futures

The omega ratio is a more general measure of risk that takes the entire distribution of an investment (e.g., a CTA) into account. Traditional measures of risks and risk-adjusted returns such as VaR, Sharpe ratio, Treynor ratio, and information ratio have been questioned by researchers because of their reliance on specific and sometimes restrictive assumptions about the distribution of returns. Omega is an alternative to these and provides a better assessment of downside risk and upside potential relative to a target return. The target level could be zero for an absolute return product such as a CTA or a hedge fund, while other rates, such as the inflation rate or a short-term nominal riskless rate, could be selected for a product that is supposed to protect the real value of an investor's portfolio.

The **omega ratio** is the ratio of the average realized return in excess of a given target return relative to the average realized loss relative to the same target return. Exhibit 26.13 contains the hypothetical monthly returns on a CTA. In this example,

EXHIBIT 26.13 Hypothetical Monthly Return (Target Rate = 4% per Year)

	Realized Monthly Return	Above Target Return	Below Target Return	Upper Partial Moment	Lower Partial Moment
Jan	18.55%	1	0	18.22%	0.00%
Feb	−18.62%	0	1	0.00%	18.95%
Mar	8.58%	1	0	8.25%	0.00%
Apr	−4.68%	0	1	0.00%	5.01%
May	1.69%	1	0	1.36%	0.00%
June	−3.50%	0	1	0.00%	3.83%
July	−13.07%	0	1	0.00%	13.40%
Aug	−21.47%	0	1	0.00%	21.80%
Sept	−4.91%	0	1	0.00%	5.24%
Oct	8.13%	1	0	7.80%	0.00%
Nov	2.66%	1	0	2.33%	0.00%
Dec	−5.27%	0	1	0.00%	5.61%
Average				3.16%	6.15%
Omega					0.51
Target					4%

a 4% annual return (0.333% monthly return) is the stated target. For each realized monthly return, it is determined whether the realized return is greater or less than the target level. These differences are presented as the upper partial moment and the lower partial moment. For example, the January figure in the upper partial moment column is given by the following:

$$18.22\% = \max(18.55\% - 0.333\%, 0)$$

and the February figure in the lower partial moment column is given as follows:

$$18.95\% = \max(0.333\% - (-18.62\%), 0)$$

The averages of these upper and lower partial moments are calculated to be 3.16% and 6.15%, respectively. Finally, the omega ratio is calculated to be 0.51, the ratio of these two figures (3.16/6.15). Mathematically, omega is given by the following expression:

$$\Omega = \frac{\text{Upper Partial Moment}}{\text{Lower Partial Moment}} = \frac{\frac{1}{N}\sum_{i=1}^{N} \max(R_i - T, 0)}{\frac{1}{N}\sum_{i=1}^{N} \max(T - R_i, 0)} \tag{26.9}$$

where R_i is the rate of return on the investment in period i, N is the total number of observations, and T is the target rate.

When the omega ratio is less than 1, it means the investment has provided fewer opportunities to earn a return that exceeds the target level. Furthermore, it can be shown that higher volatility, lower skewness, and higher kurtosis will, in general, reduce the omega of a portfolio (Shadwick and Keating (2002) and Kazemi, Schneeweis, and Gupta (2003)). In addition, increasing the target return will reduce the omega of the investment product. For instance, in the previous example, the omega will decline to 0.48 if the target return is raised to 8% per year (0.666% per month).

APPLICATION 26.3.7

Assume that the returns of an investment exceed the target rate in five periods by 1%, 3%, 4%, 5%, and 7%, while the returns fall short of the target in four periods by 4%, 6%, 8%, and 10%. What is the omega ratio?

The sums in the upper and lower partial moments are calculated to be 20% and 28%, respectively. Using Equation 26.9, and noting that the term $1/N$ cancels it out the omega ratio is simply 20%/28% = 0.71. Note that the return deviations explicitly provided in this example can be found as the absolute values of the quantity: total return – target return.

Using historical performance figures, studies have shown that for a target level of zero, the omega of a diversified portfolio of CTAs would be around 4, and the omega measure of MSCI World equities would be around 2.

Understanding investment risk in any portfolio requires constant interpretation of multiple risk factors. This is no less true in managed futures. However, doing so in managed futures is made somewhat easier because these strategies typically trade only futures contracts listed on major exchanges or liquid over-the-counter foreign exchange markets. In both cases, pricing of these instruments is transparent and continuous, thereby permitting the risk of these portfolios to be monitored and measured on a real-time basis.

26.4 THREE APPROACHES TO THE BENCHMARKING OF CTAs

In general, there are three approaches to benchmarking managed futures performance.

26.4.1 Benchmarking CTAs with Long-Only Futures Contracts

The first approach is to use an index of long-only futures contracts. Because CTAs are as likely to be short as to be long, this approach has not been found to be particularly useful. Schneeweis and Spurgin (1997) and Spurgin et al. (2011) note that there appears to be little connection between the absolute returns of major indices and the returns of CTA indices, and the results presented suggest that CTA-based indices provide a better benchmark for an actively managed futures portfolio than either passive or active long-only commodity-based performance indices.

26.4.2 Benchmarking CTAs with Peer Groups

The second approach is to use peer groups. In this case, managed futures are most commonly benchmarked to indices representing active or passive futures trading. Active benchmarks of futures trading reflect the actual performance of a universe of CTAs. As previously noted, there are a number of issues that one needs to be aware of when using hedge fund databases, including those reporting CTA performance. Recall that many funds report to only one of these providers and not others. Indeed, some of the best CTAs have no incentive to report to any database. A major problem with using an index reflecting the actual performance of CTAs is that the benchmark is not investable. For traditional asset classes, construction of investable indices is common; relatively inexpensive investable products, such as indexed mutual funds and exchange-traded funds (ETFs), are available to investors. In contrast, construction of truly investable indices in the CTA space is complex and may not result in a truly investable index that is an accurate representation of the industry.

Within the second approach, a few indices composed only of managers who invest through managed accounts at a specific firm have recently become available. This would be true of the Newedge CTA Index or the WisdomTree Managed Futures Strategy Fund. These investable indices may suffer from access bias, which could have an adverse impact on their performance. **Access bias** arises because some managers may not wish to be part of an investable index due to the restrictions that such indices may impose on them. For example, a manager may be prevented from increasing or decreasing leverage outside a predetermined level, or a manager may have to agree to reduce fees. Therefore, only those managers who have difficulty raising funds outside the investable index platform may agree to be part of the platform.

26.4.3 Benchmarking CTAs with Algorithmic Indices

Finally, CTAs may be compared with passive indices of futures trading. These passive indices represent the performance of an individual algorithmic-based trading system, as opposed to the performance of the CTAs themselves. As expected, it is impossible to create algorithmic passive indices for discretionary CTAs because, as the name implies, these managers use a significant amount of discretion, which cannot be modeled through a systematic trading approach. On the other hand, it is not difficult to develop an entirely algorithmic trading strategy to replicate the main features of a trend-following CTA. While such a benchmark may underperform most trend-following CTAs, it is likely to capture many broad features of such managers.

The oldest and most prominent of these algorithmic indices is the MLM Index. Since 1988, Mount Lucas Management has published the MLM Index, which represents the daily performance of a 252-day moving average trend-following system applied to a large set of futures markets. Such an algorithmic-based index can be quite useful because it provides the return that can be achieved from a simple trend-following trading model. This return can then be compared to the performance of active trend-following managers in order to examine the value added by more active (and expensive) strategies.

Exhibit 26.14 displays the rolling correlation between the Barclays Systematic CTA Index and the MLM Index. We can see that the correlation is generally positive and about 0.6. However, there are periods during which the correlation drops

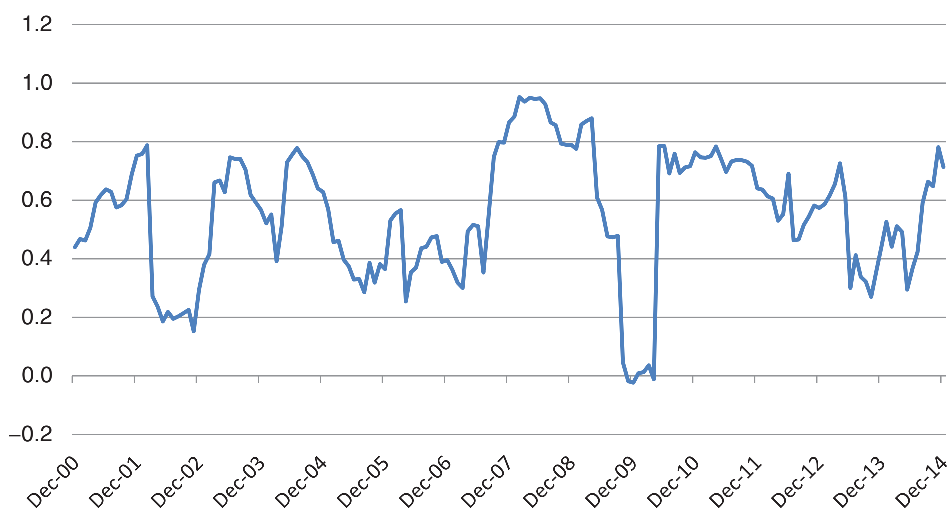


EXHIBIT 26.14 Rolling Correlation between Barclays Systematic CTA Index and MLM Index
Source: Bloomberg and authors' calculations.

close to zero. In particular, we can see that during the 2007–9 financial crisis the correlation experienced a sharp drop. This indicates that while CTA managers were relatively quick in reacting to changes in market conditions and taking advantage of opportunities that resulted from increased market stress, the MLM Index was slow to identify those changes. This may be considered good news, as CTA managers are paid to outperform simple algorithmic trading strategies. Between August 2007 and August 2009, the CTA index earned 20.8%, whereas the MLM Index was up only 12.3%. While both indices significantly outperformed equity markets, CTA managers were able to generate substantially higher crisis alpha.

26.4.4 CTA Benchmarking and Investment Strategy

Although benchmarking trend-following CTAs can be a reasonably straightforward exercise due to the high correlation among trend-following strategies, the same is not true of other managed futures trading styles, such as non-trend-following and relative value. There is considerable heterogeneity among managers in both of these styles. For example, one fund-of-funds manager noted that among 30 trend-following CTAs examined, the average paired correlation was 0.56, but among 25 non-trend-following CTAs examined, the average paired correlation was zero, indicating no common trading style among these managers. Finding or creating useful indices to benchmark non-trend-following CTA styles remains a significant challenge. A paper by Kazemi and Li (2009) examines the performance of both trend-following and non-trend-following managers using portfolios of futures contracts. The portfolios are created using the Sharpe style analysis approach, whereby returns of individual CTAs are regressed against total returns of a set of futures contracts. Kazemi and Li report that the resulting portfolios have significant explanatory power when the performance of trend-following managers is analyzed (the *r*-squareds are as high as 45%). On the other hand, the performance of non-trend-following managers can hardly be explained with an average *r*-squared of about 6%. However, non-trend-following managers displayed higher alphas in comparison to trend-following managers.

Currently, several managed futures indices publish returns on a daily basis, including the Credit Suisse (CS) Managed Futures Index, the WisdomTree Managed Futures Strategy Fund, and the Newedge CTA Index. Returns on these indices closely approximate returns on an investable product. Each index has specific distinguishing characteristics. For example, the constituent weights of the CS Managed Futures Index are calculated based on the assets under management (AUM) of each fund and rebalanced semiannually, whereas the Newedge CTA Index is an equally weighted index of CTA managers who are open to new investments.

As previously discussed, a number of index providers are now offering indices based on transparent systematic trend-following strategies. Similar to the procedure adopted in traditional asset classes, investors can use these indices to separate sources of CTA returns into beta and alpha.

26.4.5 Five Conclusions from Evidence on CTA Benchmarking

Empirical studies of the performance of these systematic trend-following strategies and their comparisons to the performance of diversified portfolios of actual trend-following CTAs lead to the following five conclusions:[3]

1. Passive trend-following indices, such as the MLM Index, can provide a reasonable benchmark for trend-following CTAs.
2. Less than half of the historical excess return earned by trend-following CTAs is due to their beta exposure to passive trend-following indices. More than half of the returns cannot be captured by such passive indices.
3. Trend-following CTAs display low exposures to traditional asset classes as well as long-only commodity indices. The same low exposure is provided by passive trend-following indices.
4. Discretionary CTAs display low exposures to traditional asset classes as well as long-only commodity and passive trend-following indices.
5. Peer groups appear to be the most suitable benchmark for discretionary CTAs.

26.5 MANAGED ACCOUNTS AND PLATFORMS

Once investors decide to make an allocation to managed futures, they must tackle the problem of just how to structure the investment. The choice of vehicle employed to make an allocation is dependent on the size of the allocation as well as on the level of expertise and experience the investor possesses.

26.5.1 Factors Affecting the Choice of Investment Product

To tackle the problem of structuring the investments, investors will need to follow a decision-making process that proceeds along the following lines.

First, the investor must determine how many CTAs he wants in the portfolio. Many family offices, and even some larger institutional investors, will decide to invest in a single CTA. This decision has the virtue of simplicity and is possible to implement by choosing one of the large, diversified trend-following CTAs, whose performance correlates highly with a trend-following benchmark. If the investor decides to use this approach, the focus should be on examining differences between investing in a fund sponsored on behalf of the CTA versus a managed account.

Generally, the single-CTA route exposes the investor to a greater amount of risk. In this scenario, the results depend on the performance of a single manager, concentrate risk to a single organization, and may be exposed to a limited number of trading models. To avoid these constraints, the investor may decide to form a diversified portfolio of CTAs.

Second, if the investors are large enough, they must decide how to create a diversified portfolio of CTAs. Initially, the most cost-effective approach to achieving diversification is to allocate to a multi-CTA fund. Then, as the size of the allocation to CTAs increases, more options become available to the investor.

26.5.2 Four Factors Driving CTA Portfolio Construction

Eventually, the investor must decide whether to assemble an in-house team to manage the portfolio and whether to use a managed account platform. There are cost issues associated with each choice. Related to the size of allocation to CTAs, two issues arise as the investor decides on the best approach to creating a diversified portfolio of CTAs.

The first issue is related to the level of allocation at which it becomes cost-effective to move from working with a multi-CTA investment program to assembling an in-house team to create and manage a diversified portfolio of CTAs. While no exact figure exists for this cost threshold, it is primarily affected by four factors:

1. The extra layer of fees that the investor will have to pay the multi-CTA fund
2. The cost of assembling a team of analysts who can construct and manage a CTA portfolio
3. The minimum size of the investment that CTAs are willing to accept
4. The number of CTAs that should be included in the portfolio to achieve diversification

For example, in order to achieve a reasonable degree of diversification, a portfolio may consist of about six CTAs. If the minimum investment size for large institutional-quality funds is assumed to be $5 million, an investment of at least $30 million is required to make it cost-effective to create a portfolio of CTAs. The management fee associated with a $30 million portfolio is between $300,000 and $450,000. This amount may not be enough to create a team that can select and manage a portfolio of CTAs. As the size of the investment increases, the extra level of fees paid increases as well, and it will eventually become economical to create an in-house team to manage a portfolio of CTAs and forgo allocations to multi-CTA funds.

The second issue is related to the next level of allocation when it becomes viable for the investor to use a managed account platform. Managed accounts offer a number of very important advantages over CTA funds, including transparency, security of collateral, and ease of opening and closing positions. However, they require the investor to have experienced people and reliable systems in place, which can be costly.

26.5.3 Structure of CTA Products and CTA Funds

A CTA fund is a separate legal corporation with a board of directors that assigns trading authority to the single-manager CTA's investment management company. With a CTA fund, the investor turns the money over to the company organized by the CTA manager by filling out a subscription form and wiring the money to the account. Investors have limited liability for the fund manager's actions. Some oversight is provided by the board of directors and third parties, such as auditors, brokers, and regulators, because of the fund's status as an independent company. Additionally, for the manager's flagship fund, industry analysts, reporters, and other investors in the fund would pay attention to performance, effectively providing additional oversight.

Fund investors have a broad choice of managers, because most managers have funds, and funds are more likely to accommodate smaller investors than are managed accounts. Managers like dealing with investors through funds because there are standard terms for risk, return, and liquidity. However, a fund has less transparency for an investor than does a managed account, and the amount of transparency can be variable. It is customary to provide monthly returns and a performance commentary. Liquidity can be just as variable, though, with managers offering different schedules for withdrawals, ranging from daily to annual, and requiring notice in advance of the scheduled withdrawal period.

These accounts generally use a master-feeder structure. Under a typical master-feeder arrangement, two funds feed into a single master account. For example, one feeder could be for onshore (e.g., U.S.) investors, while the other is for overseas investors. Money from both feeder funds is combined in the master fund, which opens trading accounts at the brokerage firm. This setup is established for tax purposes, enabling overseas investors to avoid dealing with onshore tax authorities, while onshore investors receive the documentation they need to report to the local tax authorities.

Some funds are structured with different share classes, which may differ in fee structure and withdrawal rights. For example, someone who wants daily liquidity might be willing to pay a higher management fee than someone willing to accept annual liquidity. A new investment may be held in a temporary share class until it reaches the same high-water mark as the rest of the fund. Some classes may be invested in additional assets that are not part of the main fund.

Multi-CTA funds are known variously as CTA funds of funds or commodity pools. From the investor's standpoint, both accomplish the same thing: They provide a single vehicle for investing in a diversified portfolio of CTAs. The differences are chiefly regulatory, relating to the way the funds are structured and where they are offered. Commodity pool operators (CPOs), for example, create vehicles that are distributed in the United States. They are common investment vehicles for retail investors, high-net-worth individuals, and even some small institutions. Some funds launch offshore funds and tend to attract larger institutional investors. The expression *fund of funds* derives from a time when the primary investment vehicle at the individual CTA level was a fund. As the industry has evolved, multimanager funds have migrated to the use of managed accounts.

The primary benefits of a multi-CTA fund structure are accessing the expertise that the fund manager has in choosing the managers, structuring the portfolio, performing both investment and operational due diligence, reporting performance, monitoring risk, and accounting. In addition, the fund manager performs less obvious tasks, such as collecting data, meeting managers, running background checks, analyzing performance and strategies, negotiating contracts and fees, monitoring performance, and rebalancing the portfolio as necessary. From an investor's perspective, the investment offering and services of such a structure consolidate much of the work into choosing and reviewing a single organization. Because individual multi-CTA funds have different investment objectives, investors need to find a fund that is consistent with their needs for risk and return as well as for reporting and transparency.

With a multi-CTA fund, the manager assembles a portfolio of CTAs and then accepts investments in the entire portfolio. The multi-CTA manager charges a fee for portfolio construction and oversight services. Each of the managers in the fund charges a fee as well. Although investors negotiate contracts and fees with the multi-CTA manager directly, and the investment is consolidated into a single organization, investors still have due diligence and monitoring obligations.

The fees charged by multi-CTA managers raise important questions about how to structure the investment. If the level of allocation is relatively small, the investor would likely invest in CTA funds. However, if investors intend to allocate a large amount to CTAs, they might be better advised to save the fees that would be paid to a multi-CTA manager and simply hire the staff and consultants needed to select CTAs, perform the due diligence, construct the portfolios, and so on. Managed accounts, which are the vehicle of choice for most multi-CTA managers, are a much bigger undertaking than are CTA funds. To warrant the work involved in setting up brokerage accounts, negotiating agreements, monitoring the accounts, reconciling trades, complying with anti-money-laundering regulations, managing cash flows, and so forth, a reasonable break-even point is an investment in CTAs of around $500 million.

26.5.4 Structuring CTA Products with Managed Accounts

A managed account is a brokerage account, held by a brokerage firm that is also registered as a futures commission merchant, in which investment discretion has been assigned to the CTA manager. The investor is responsible for opening and maintaining the account, reconciling brokerage statements, and maintaining cash controls, as well as negotiating contracts with managers, including investment management agreements and powers of attorney. The limited power of attorney gives the manager authority to trade on the investor's behalf, but the money has to remain in the investor's account. The investor controls the terms of the power of attorney, including the right to revoke trading privileges.

The key advantage to a managed account is complete control. By pulling trading privileges, the investor has the ability to manage the cash and liquidate the account at any time. Managed accounts, then, avoid the lockup and gating provisions frequently found in hedge fund investments. In theory, this gives the investor better than daily liquidity, as the account can be liquidated whenever the market is open. That alone is enough to make some investors demand managed accounts, especially investors with in-house staff to handle the paperwork.

Managed accounts have other advantages. The money is within the investor's control, not the fund manager's, at all times. The accounts offer complete transparency. The investor can see the positions, trades, and details at any time. Managed accounts, then, virtually eliminate the risk of fraud, as the transparency and security of these accounts prevent the manager from misstating leverage, manipulating returns, or stealing the investor's assets. The investor can choose the parameters for leverage based on the targeted volatility of returns. The choice of leverage makes it easier for the investor to manage the underlying cash. In fact, this type of CTA account structure is often looked at as an overlay on the cash position in an investor's portfolio, rather than a separate asset class.

Of course, these advantages come at a cost. The first is the reduced pool of managers to choose from. Many large managers do not accept managed accounts, whereas those that do require a large minimum investment and other administrative stipulations. In addition, the previously mentioned transparency and control come with the responsibility for establishing and maintaining brokerage accounts that require legal, administrative, risk, and investment oversight in accordance with each organization's investment standards. Further, unless procured by the investor, there is no administrator or auditor.

Managed accounts can be set up in a variety of ways to meet different portfolio policy requirements. The limited partnership structure of hedge funds limits investor liability to the amount invested. For example, an investor who allocates $10 million to a failed hedge fund cannot lose more than $10 million, even if the hedge fund is highly leveraged and sustains losses greater than the amount of contributed client assets. Managed accounts, however, do not automatically have a limited liability structure. Especially in futures markets, where the required margin is much smaller than the notional value of contracts, investor losses in high margin-to-equity investments can be larger than the amount of contributed capital. Therefore, managed accounts must be carefully designed with a legal structure that ensures that limited liability is obtained. Structures offering limited liability vary by legal jurisdiction but may include limited liability companies, limited partnerships, special purpose vehicles (SPVs), or bankruptcy-remote entities. Each structure is designed to limit investor losses to the amount of cash invested, even if trading activity incurs greater losses.

In many managed account situations, the investor begins by setting up an SPV or another holding entity to fence off any trading liabilities from the rest of the money that the investor controls. It is not a necessary step, however, as there are other ways to manage the potential liability. In most case, the investor uses the SPV to open an account at a brokerage firm where the CTA manager has trading authority. The investor gives the manager the authority to trade in the account.

26.5.5 Structuring CTA Products with Platforms

An alternative way to structure a CTA investment is through a platform. This is a relatively new product, offered by a handful of financial services firms. It operates almost like a multi-CTA fund, except that investors can select their own leverage and create their own portfolios from the mix of funds offered through the platform.

Platform companies argue that a key advantage of their structure is having objective, independent boards of directors and vendors that are selected by the platform company, not a manager. The platform structure may also reduce custody concerns. Usually, these platforms pass on some of the advantages of managed accounts, such as transparency, liquidity, and customized leverage.

Investors can have a series of fund investments in the platform's participating money managers, receiving consolidated performance information as well as consolidated subscription and redemption paperwork from the platform. It is relatively easy to move money from one manager to another. Because of the transparency and liquidity, these are a hybrid of managed accounts and CTA funds.

Exhibit 26.15 consolidates much of the previous discussion into a summary table, providing an overview of the primary characteristics of the four types of investment structures: CTA funds, multi-CTA funds, managed accounts, and platforms. While there are definitely exceptions to the assignment of characteristics in this table, it should serve as a good starting point when considering an investment structure.

26.6 CONCLUSION

Available evidence suggests that CTAs benefit investors by providing meaningful diversification benefits to portfolios consisting of traditional asset classes as well as hedge funds. To achieve these benefits, investors need to consider the various ways that CTAs can be accessed. Further, to mitigate risks associated with investing in a single manager (e.g., model risk, key-person risk, and firm risk), the investor should consider investing in a portfolio of CTAs. Another consideration is whether to pay an additional layer of fees in order to outsource the construction of this portfolio to

EXHIBIT 26.15 Structural Characteristics of CTA Funds, Multi-CTA Funds, Managed Accounts, and Platforms

	Liability	Liquidity	Funding and Leverage	Oversight and Control of Assets	Maintenance	Position and Trade Transparency	Availability	Due Diligence Burden
CTA fund	Limited	Monthly	Manager determined	Directors selected by manager	Low	Usually not	Most managers offer flagship fund.	Medium
Multi-CTA fund	Limited	Weekly/monthly	Manager determined	Directors selected by manager	Low	Usually not	Most managers offer flagship fund.	Low
Managed account	Unlimited	Daily	Customer determined	Investor	High	Yes	Not all managers accept managed accounts.	High
Platform	Limited	Weekly/monthly	Hybrid	Directors selected by platform	Medium	Varies; manager determined	Not all managers have an established relationship with a platform.	Low

Source: Newedge Prime Brokerage Research.

a multi-CTA fund, or to assemble an in-house team that could select and construct the portfolio. Finally, for those investors who plan to make a substantial allocation to CTAs, a managed account platform will provide substantial transparency, control, and customization of the CTA investment program.

To understand the risk-return properties of CTAs, one needs to apply certain measures designed to account for the special characteristics of CTAs. Two facts, (1) that CTAs allow investors to gain exposure to a variety of markets with minimal investments, and (2) that the margin requirement for each contract is closely linked to the riskiness of the underlying market, mean that investors should carefully examine the amount of margin that a CTA is carrying. In addition to the margin-to-equity ratio, investors should use other metrics—such as volatility, VaR, and CaR—to gain a better understanding of a CTA's risk profile.

One important issue that investors need to address is the lack of proper benchmarks for CTAs in general, and discretionary CTAs in particular. Although passive strategies have been developed to mimic the risk-return properties of trend-following CTAs, no such passive strategies have been developed for discretionary CTAs. The MLM Index represents one such attempt to produce a passive and investable benchmark for systematic CTAs. The performance of this index can be used to estimate the returns on CTAs due to their beta exposures to a passive trend-following strategy. The results presented in this chapter show that less than 50% of the return earned by trend-following CTAs is due to their beta exposure to a passive trend-following index.

This chapter has also examined the role of CTAs in providing downside protection for traditional asset classes such as equities and high-yield bonds, as well as alternative assets such as hedge funds. It was argued that although CTA return profiles give the appearance of being long volatility, it is more appropriate to characterize them as being long gamma, as trend-following CTAs increase the delta of their positions as prices move in their favor, the basic characteristic of a long gamma position.

NOTES

1. In this example, crisis alpha is defined as the average monthly return of each strategy/index during the months when the MSCI World Index's return was one standard deviation or more below its mean.
2. Futures contracts have zero value at inception, requiring no capital to change hands. Only good-faith margin must be posted to cover short-term losses in the position, which are settled on a daily basis.
3. See Schneeweis and Kazemi (2014).

REFERENCES

Edwards, F., and J. Park. 1996. "Do Managed Futures Make Good Investments?" *Journal of Futures Markets* 16 (5): 475–517.

Elton, E., M. Gruber, and J. Rentzler. 1987. "Professionally Managed, Publicly Traded, Commodity Funds." *Journal of Business* 60 (2): 175–99.

Fung, W., and D. A. Hsieh. 1997. "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds." *Review of Financial Studies* 10 (2): 275–302.

Greyserman, G., and K. Kaminski. 2014. *Trend Following with Managed Futures: The Search for Crisis Alpha*. Hoboken, NJ: John Wiley & Sons.

Kaminski, K. 2011. "Regulators' Unintentional Effects on Markets." *SFO*, July.

Kat, H. 2002. "Managed Futures and Hedge Funds: A Match Made in Heaven." Research paper, Cass Business School, London.

Kazemi, H., and Y. Li. 2009. "Market Timing of CTAs: An Examination of Systematic CTAs vs. Discretionary CTAs." *Journal of Futures Markets* 29 (11): 1067–99.

Kazemi, H., T. Schneeweis, and R. Gupta. 2003. "Omega as a Performance Measure." Working paper, Isenberg School of Management, University of Massachusetts.

Malek, M. H., and S. Dobrovolsky. 2009. "Volatility Exposure of CTA Programs and Other Hedge Fund Strategies." *Journal of Alternative Investments* 11 (4): 68–89.

McCarthy, D., T. Schneeweis, and R. Spurgin. 1996. "Investment through CTAs: An Alternative Managed Futures Investment." *Journal of Derivatives* 3 (4): 36–47.

Schneeweis, T., and H. Kazemi. 2014. "Benefits of Managed Futures." Center for International Securities and Derivatives Markets, Isenberg School of Management, University of Massachusetts.

Schneeweis, T., and R. Spurgin. 1997. "Comparison of Commodity and Managed Futures Benchmark Indices." *Journal of Derivatives* 4 (4): 33–50.

Shadwick, W. F., and C. Keating. 2002. "A Universal Performance Measure." *Journal of Performance Measurement* 6 (3): 59–84.

Spurgin, R., T. Schneeweis, and G. Georgiev. 2001. "Benchmarking Commodity Trading Advisor Performance with a Passive Futures-Based Index." CISDM Working Paper.

CHAPTER 27

Relative Value Strategies

Rather than being a specific strategy, relative value refers to any strategy that attempts to benefit from relative mispricing of two or more securities. For the purpose of this chapter, a strategy is referred to as being relative value if its typical implementation requires one to take long and short positions in instruments that are linked. This linkage could be based on an economic relationship between the two securities or on empirical observations. For example, stock prices of two oil companies are likely to be highly linked and, in the absence of major firm-specific events, will tend to follow the same path. Likewise, the stock prices of a manufacturer and its suppliers are linked and thus will tend to follow similar paths. Finally, empirical studies have shown there exists a negative relationship between equity returns and estimates of their volatility; that is, negative returns are associated with increased return volatility. If the relationship between two securities or economic variables is strong and stable enough, then it could be the basis of an arbitrage or a relative value strategy.

This chapter begins with a discussion of limits to arbitrage, which attempts to explain why apparent arbitrage opportunities may not be fully taken advantage of by money managers. It argues that when traders have to raise funds to implement these strategies, they may allocate only a portion of their capital to available arbitrage opportunities; therefore, such opportunities may not disappear as quickly as one would expect.

The chapter then discusses two arbitrage or relative value strategies: convertible bond arbitrage and pairs trading. Although the total amount of assets managed under these strategies may not be significant, discussion of them is useful in highlighting various aspects of relative value and arbitrage strategies.

27.1 LIMITS TO ARBITRAGE OF RELATIVE VALUATION

The textbook definition of *arbitrage* refers to a situation in which an investor is able to earn riskless positive profits while making no or very little investment. The simplest example is when the same asset is trading at different prices in two separate markets, and the cost of trading in these markets is rather negligible. In this case, a trader would short the asset in the market where the price is relatively high and use the proceeds to buy the same asset in the market where the price is lower. Assuming that the prices would eventually converge, the trader is expected to earn a riskless positive profit with no investment. Of course, such an easily exploitable arbitrage opportunity is not likely to arise and, if it does, is not likely to last for more than a few seconds.

27.1.1 Absence of Arbitrage in Derivative Pricing

Almost the entire modern finance literature is concerned with the relationship between risk and return and how investors are compensated for taking more risk. The efficient market hypothesis (EMH), which was discussed in CAIA Level I, is founded on the idea that competitive markets almost never offer investors free lunches, meaning that investors should not expect to earn large positive returns without taking higher levels of risk. Modern portfolio theory has been developed to let investors know how to diversify the unnecessary risks of their investments so that they bear only those risks that are rewarded by the marketplace. Therefore, the idea that arbitrage opportunities may arise and may last for relatively long periods of time appears to contradict much of what is discussed in finance books.

Absence of arbitrage is the cornerstone of derivative pricing models such as the Black-Scholes option pricing model and the cost-of-carry model for the futures markets. Consider the following example, which illustrates the put-call parity relationship and begins with two portfolios:

- Portfolio A: A European call option with maturity date T and strike price K, plus a zero-coupon bond that pays K at time T
- Portfolio B: A European put option on the same underlying asset as the call option, with the same maturity date and strike price, plus one share of the underlying stock

Now consider the payoffs from these two portfolios under two mutually exclusive states of nature: $S_T < K$ and $S_T \geq K$, where S_T is the stock price at time T (see Exhibit 27.1).

Both portfolios are worth the same at time T, which means they should be worth the same today; otherwise, arbitrageurs would take advantage of the mispricing. It turns out that one rarely sees violations of put-call parity that are large enough to exceed the transaction costs associated with the strategy. However, as the next section discusses, there are circumstances under which apparent arbitrage opportunities or at least exceedingly good deals may remain available for more than a few seconds.[1]

EXHIBIT 27.1 Put-Call Parity

		$S_T \geq K$	$S_T < K$
Portfolio A	Call option	$S_T - K$	0
	Zero-coupon bond	K	K
	Total	S_T	K
Portfolio B	Put option	0	$K - S_T$
	Share	S_T	S_T
	Total	S_T	K

27.1.2 Examples of Arbitrage Opportunities

In theory, arbitrage opportunities are short-lived, and successful arbitrageurs will profit from their transactions while steering security prices toward their fundamental values. **Fundamental value**, also known as intrinsic value, refers to the value of an investment based on a comprehensive analysis of underlying economic factors and forces. Hence, arbitrageurs are integral to the working of an efficient market and add significant economic value to the price discovery process. In a highly efficient market, profit making from such arbitrage opportunities occurs quickly, and hence the speed of convergence to efficient prices is often touted as a measure of market efficiency.

Consider the following examples illustrating arbitrage opportunities:

1. A local community bank charges a 4% annual lending rate; a neighboring local bank pays 5% per year for deposits. Spell out the arbitrage opportunity.
2. Explain the arbitrage opportunity when the price of a dually listed mining company stock is USD 50 on the New York Stock Exchange and CAD 52 on the Toronto Stock Exchange. Assume the exchange rate is USD 1 = CAD 1.01.
3. XYZ Corporation owns 80% of ABC Corporation. Both have 100 million shares outstanding. Shares of XYZ Corporation are trading at $90 a share, while shares of ABC Corporation are trading at $120 a share. Is there an arbitrage opportunity?

How might arbitrage be performed in these three scenarios? The following list demonstrates methods of arbitrage corresponding to each of the three scenarios:

1. Borrow any sum for one year from the bank charging a 4% lending rate and deposit the same amount in the bank offering a 5% return on deposits for the same term. At the end of the year, arbitrage profit amounts to (Borrowed Sum) × (5% – 4%). This arbitrage opportunity seems too easy to become available. In fact, if there is no limit on the amount that can be borrowed, the investor could generate an unlimited amount of profit.
2. Take a long position in the stock in New York for $50, and sell the stock in Toronto for $51.485 (CAD 52). Note that it is assumed that shares listed in New York cannot be sold in Toronto. Again, such an opportunity is unlikely to appear. However, even it becomes available, a trader may not be willing to allocate all of her capital to take advantage of this opportunity. The reason is that the trader will earn a profit of $1.485 per share from this strategy only if the two prices converge, or if one share of stock can be freely exchanged for another. Suppose a large U.S. fund is liquidating its large position in the stock, causing its price in New York to further diverge from its price in Toronto. Will the trader be able to raise additional funds to cover her temporary losses?
3. Since XYZ owns 80% of ABC, its total market value must be at least 80% of the total market value of ABC. This means its shares must sell for at least $96 a share. An arbitrageur will short shares of ABC and will take a long position in the same number of shares of XYZ. Once the market value of XYZ converges to 80% of the market value of ABC, the trader would benefit. For example, share prices of both companies may increase while they are converging. Suppose shares of ABC rise to $200 per share while shares of XYZ rise to $190 share.

This means that as a stand-alone corporation, XYZ is valued at $30 per share. The trader has earned $10 in profit per share. What if the owners of the shares that were borrowed in order to short ABC's stock demand to have them back before share prices have converged? What if the prices drift even further apart, and thus the trader is forced to cover his short positions? It may not be wise for this trader to allocate his entire capital to this strategy.

These examples show that engaging in arbitrage may expose traders to certain risks. While a trader is likely to earn arbitrage profits, the interim risks may be large enough to convince the trader to allocate only a small portion of her funds to the strategy. Almost all potentially available arbitrage opportunities involve some degree of risk, at least in the interim. The term **risk arbitrage** is often used by the investment industry to describe profit opportunities that involve some limited risk but have the potential to provide large positive returns relative to the amount of risk involved. Risk arbitrage contrasts with pure arbitrage, which describes riskless arbitrage opportunities.

27.1.3 Factors Affecting Limits to Arbitrage

A number of factors limit pure and/or risk arbitrage activities. These factors are typically ignored in the textbook definitions of arbitrage, while in real-world cases they limit the amount of capital traders are willing or able to commit to some arbitrage opportunities. As a result, certain arbitrage or risk arbitrage opportunities may persist through time. There are a number of risks and other factors that contribute to limits to arbitrage:

- Fundamental risk: **Fundamental risk** emanates from an unexpected change in the fundamental value of a security, causing an apparent arbitrage opportunity to generate losses on the part of the investor. Suppose the same commodity is produced and traded in two countries. A trader observes the gap between the two prices of the commodity and determines that it is much wider than its historical range. Given that there is free trade in the commodity, the trader believes the gap will approach its historical value and therefore takes a long position where the commodity is selling at a low price and a short position where the price is high. However, the country where the commodity is trading at a high price unexpectedly decides to impose tariffs on importation of the commodity. In this case, there has been a fundamental change in the economic relationship between the two commodities, and the trader is likely to face some losses. Of course, experienced traders will understand the risk of such an unlikely event and will therefore commit a relatively small amount of capital to the strategy. As a result, even in the absence of tariffs, the two prices may not converge immediately.
- Noise traders: The textbook description of the efficient market hypothesis assumes that investors are rational and act only on information that can affect the fundamental value of securities. The term **noise traders** refers to those investors who trade securities for reasons not related to the fundamental value of securities. For instance, some investors may sell securities in order to meet liquidity needs, while others may decide to buy securities simply because their prices have been increasing in recent days. That is, they follow a momentum strategy. Recent

developments in the field of behavioral finance have identified a number of biases that impact investors' behavior and lead them to act as noise traders:

- Overconfidence: Studies in psychology have demonstrated that traders systematically overestimate their ability to value firms and make price and earnings predictions. They become overly bullish about stocks they are optimistic about, or overly bearish with respect to stocks they regard with pessimism.
- Representativeness: The representativeness bias is a decision-making shortcut that uses past experiences to guide the decision-making process. The term *representativeness* indicates that when investors are confronted with a new experience and need to make a judgment or decision about that situation, their brains automatically rely on past experiences and mental representations that seem similar to this new situation in an effort to guide their judgments and decisions. For example, investors who have seen Internet companies perform well in the past may consider any dot-com company to be a fast-growing firm. In other words, the person uses categories to decide if a firm will be growing quickly going forward.
- Anchoring: This term describes the common human tendency to rely too heavily on the first piece of information offered (the "anchor") when making decisions. In particular, this arises when a person is supposed to estimate a number. During decision making under uncertain circumstances, anchoring occurs when individuals use an initial piece of information to make subsequent judgments. In the stock market, anchoring may lead investors to believe that a stock that was trading for $100 a few weeks earlier and is now trading for $20 must be cheap, and therefore the investor may decide to buy the stock even though the fundamental value of the stock may now be $20 or lower.
- Disposition to loss aversion: Empirical and laboratory evidence demonstrates that people have stronger reactions to losses than to gains. In addition, it is observed that investors are prone to holding on to losing stocks too long while realizing gains too quickly. Thus, even if there is a negative fundamental change in the value of a stock, investors who have suffered losses will be slow to realize those losses, causing the stock to adjust to negative news more slowly.

- Leverage risk: The profit margins associated with some arbitrage opportunities may be too small to make them a viable investment unless the investor can use substantial leverage. **Leverage risk** arises because a fund manager who is using leverage may suffer temporary losses, and as a result, the lenders decide to reduce or eliminate the fund's line of credit. Consequently, the fund is forced to liquidate some of its positions at a loss, leading outside investors to withdraw their capital from the fund.
- Market frictions: **Market frictions** are impediments to costless trading—such as transaction costs, taxes, and regulations—that can create market imperfections and make it too costly or too risky to implement certain arbitrage strategies. The textbook definition of arbitrage assumes that markets are perfect. For example, it is generally suggested that there are no barriers to shorting overvalued securities. However, in the real world, it may be costly or even impossible to short certain securities; also, investors who have short positions might be forced to cover their positions if the owners of shares demand them back. The term *short squeeze* refers to a situation in which investors with short positions are

forced to cover their positions because of an increase in the price of the shorted security.

- Agency relationship: An agency relationship (i.e., a principal–agent relationship) arises whenever an owner of an asset hires an agent to manage that asset on the owner's behalf. This relationship gives rise to certain problems and costs. For example, how would the owner know that the agent is managing the asset for the sole benefit of the owner, or that the agent is not taking too much risk? The owner may have to hire auditors and install monitoring systems to ensure that the assets are properly managed. This will be costly. One implication of an agency relationship is that in order to reduce the chance of underperformance relative to their peers, agents may decide to follow the crowd rather than invest in unique strategies. Also, there is a risk that the owner may not be patient enough to stay with certain strategies. Suppose an agent identifies a highly undervalued security. For the price of this security to reach its potential fundamental value, other market participants must be convinced that the security is indeed undervalued. This could take many months, and in the meantime the security may decline even further. The owners of the asset may look at the temporary losses, conclude that the trader lacks the necessary skills, and decide to withdraw their funds.

Much of the preceding discussion can be summed up by John Maynard Keynes's statement that "markets can remain irrational a lot longer than you and I can remain solvent." This means that for a number of reasons, certain arbitrage and risk arbitrage opportunities may appear and persist for relatively long periods. Those traders who can manage and withstand the risks associated with irrational markets and noise traders should be able to take advantage of these opportunities and generate abnormal returns (net of fees) for their investors.

27.1.4 Efficiently Inefficient Markets

As previously noted, the efficiency of market prices is one of the central concepts in finance. The central thesis is that security markets are perfectly efficient and that prices reflect all available information at all times. However, this leads to two paradoxes: First, if no one will have an incentive to collect information in an efficient market, how does the market become efficient? Second, if asset markets are efficient, then positive fees paid to active managers imply inefficient markets for asset management. In other words, one cannot simultaneously assume that financial markets are dominated by rational investors who arbitrage away pricing inefficiencies and that there are irrational people who invest with professional money managers who, according to the efficient market hypothesis (EMH), do not add any value. Why should professional money managers exist at all, and why should some investors be willing to pay them substantial fees to manage their assets? As discussed in the previous section, there are many reasons to believe that arbitrage opportunities may appear in markets and that these opportunities may not be arbitraged away very quickly.

The presence of active money managers provides support for this idea. The presence of these managers seems to imply that markets are not perfectly efficient, as thousands of hedge funds, private equity funds, and active mutual funds earn substantial fees for managing investors' assets when, according to the EMH, they should underperform passive strategies on an after-fee basis. However, markets cannot be highly inefficient, because if that were the case, then many more people would enter the active money management business to earn a portion of those fees and, in the process, help make markets more efficient. Therefore, it seems that there must be a fine balance between efficiency and inefficiency. Under these circumstances, markets are said to be efficiently inefficient.

Efficiently inefficient presents the idea that markets are, on average, just inefficient enough to compensate managers and investors for the costs and risks of pursuing skill-based strategies, but not too inefficient to present a large number of money managers with easy-to-exploit arbitrage opportunities.[2] Therefore, the flow of capital to actively managed strategies is limited in a world that is efficiently inefficient. In such a world, competition among active money managers will result in markets that are almost efficient; however, inefficiencies do exist that reward those who can identify and exploit them.

It can be argued that professional asset managers arise naturally as a result of the returns to scale in collecting and trading on information. Active managers collect information about securities and then invest using this information to benefit others. Therefore, active asset managers are central to understanding market efficiency. In a market characterized as being efficiently inefficient, there exist a limited number of market inefficiencies that can be exploited by some money managers. However, finding the right manager takes time and resources, and therefore investors have to decide whether to spend search costs to find an active asset manager or allocate their capital to a passive strategy. At the margin, investors become indifferent between passive investing and searching for an active asset manager. If search costs are low such that investors can easily identify good managers, then more money is allocated to active management and some pricing inefficiencies are arbitraged away.

The remaining sections of this chapter examine two relative value strategies. The first is convertible arbitrage, which involves long and short positions in convertible bonds and equity of the same firm. The second is pairs trading, which is a quantitative market-neutral strategy trading long and short positions in equities of two firms. Pairs trading is considered one of the strategies employed within an equity market-neutral strategy. Essentially, these two strategies exemplify a range of investment options combining securities and payoffs that are subject to mispricing and arbitrage. In convertible arbitrage, a hedge fund manager exploits inefficiency between two securities from the same issuer—convertible bonds and stocks—whereas in pairs trading, a manager includes long and short positions in over- and underpriced stocks of two firms.

27.2 CONVERTIBLE ARBITRAGE: AN OVERVIEW

In its most rudimentary form, convertible arbitrage is a form of relative value arbitrage that attempts to exploit relative mispricing of convertible bonds and equities of the same issuer.[3]

27.2.1 Three Broad Steps to Convertible Arbitrage

A typical convertible arbitrage strategy entails buying underpriced convertible bonds, which allows bondholders to convert their holdings to equity at an agreed conversion ratio while selling short the underlying equity to create a hedged market-neutral portfolio. To implement this strategy, a manager follows these three broad steps:

1. Valuing the convertible bond to determine if the issue is indeed underpriced
2. Using the valuation model to determine the appropriate hedge ratio, which will determine the size of the short position in the underlying stock
3. Managing the portfolio's risks, which may include credit, interest rate, event, and liquidity risks

Three distinct issues related to convertible arbitrage are discussed in the following paragraphs. First, why are convertible bonds underpriced, and how is the extent of underpricing in a convertible bond determined, given the characteristics of the convertible bond and its underlying stock price? Second, what are the trading strategies that can take advantage of the mispricing of convertible bonds? And third, what are the explicit and implicit risks in a convertible arbitrage strategy?

27.2.2 Convertible Bond Underpricing

There are economic reasons as to why convertible bonds may be underpriced. A straight bond promises two distinct cash flows: a fixed periodic coupon until maturity, and the face value, which is paid to bondholders at maturity. The price of the bond is the discounted value of those two cash flows where the discount rate captures the risks associated with the cash flows. Unlike straight bonds, convertible bonds offer the holders the right to convert the bond to equity. Therefore, a convertible bond contains an embedded call option on the underlying common stock; hence, its value is affected by variables that determine the value of a call option—namely, the stock price and its volatility. As shown in Exhibit 27.2, holding the discount rate constant, the convertible bond price is a nonlinear function of the price of the underlying stock.

In Exhibit 27.2, the *x*-axis denotes the stock prices, while the *y*-axis denotes the theoretical value of the respective investment in bonds, stocks, and convertible bonds

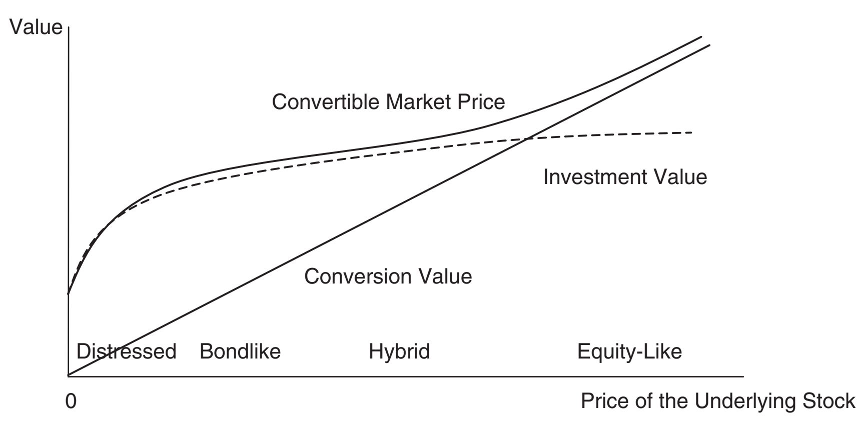


EXHIBIT 27.2 Convertible Bond Price and Price of Underlying Stock

with respect to those stock prices. The value of investment in a straight bond and stock are given by the almost horizontal dashed line and the 45° solid line, respectively; the conversion value—the theoretical price of a convertible bond—is a nonlinear function indicating conversion and investment premiums, respectively, to the left and right of the intersection between the horizontal (bond value) and the 45° (stock value) lines. Note that a convertible bond is a hybrid security that mirrors the value of a bond for low values of the stock and behaves more like a stock for high values of the stock. This stems from the fact that the theoretical payoff from a convertible bond is similar to that of a portfolio consisting of a call option and a straight bond.

The difficulty in valuing an option arises from the fact that its future payoff is contingent on the future price of the underlying stock, and any long-run price forecast for stocks is bound to be extremely noisy. It is known from option pricing models that return volatility of the underlying asset is the major determinant of the option price. Since volatility is not directly observed, the uncertainty associated with the actual return volatility of the underlying asset and its future path may lead to mispricing of the option embedded in a convertible bond.

The underpricing of convertible bonds is supported by both empirical observations and economic rationale. Coupons and yields are lower on convertible bonds than on straight bonds, as investors ascribe value to the embedded equity call option. Convertible bond underpricing is an empirical regularity, but there is also a strong economic reason for why the issuers may strategically underprice convertible bond issues.[4] So, why do corporations issue underpriced securities?

The majority of convertible bond issues are below investment grade. Therefore, for firms that typically issue convertible bonds, raising capital through straight bonds may prove too expensive. Straight bonds are likely to have a high yield, and if the bond is not a deep-discount bond, it will have to carry a high coupon rate as well. This would adversely affect the cash flows of the firm. Alternatively, these firms could issue equity, which would not have the same negative cash flow impact. However, for firms with a low credit rating, issuing equity is not an attractive option. First, the size of the issue might have to be so large that it would dilute the ownership of current shareholders. Second, stock prices of firms that issue new equity usually react negatively to such an announcement. The primary reason for this is the adverse selection problem. Equity investors conclude that, on average, only firms with a negative outlook would be willing to issue new shares and bring in new shareholders. While markets may react negatively to issuance of convertible bonds, the size of the reaction tends to be smaller.

From the issuer's perspective, the underpricing of the convertible bond is necessary to attract investors, but it also helps the firm secure equity capital at a relatively low issuance cost and avoid possible underpricing associated with seasoned equity offerings. Further, in terms of timing of the issue, the firm benefits from knowing that the bonds will be converted into shares of common stock when the stock price is relatively high, possibly overpriced, meaning that the firm may never have to pay back the principal of the bond. From the investors' perspective, this intuition that convertible bonds may be underpriced relative to the current value of the firm's stock price is the primary rationale behind a convertible arbitrage strategy. In summary, convertible arbitrage managers provide a service to issuing corporations and get paid for this service. This provides a partial economic explanation for the potential source of alpha found within the convertible arbitrage strategy.

27.2.3 Arbitrage with Convertible Bonds

This section focuses on the most common and basic convertible arbitrage strategy, in which the manager takes a long position in the convertible bond and then shorts an appropriate amount of the underlying stock in order to create a delta-hedged portfolio. That is, in the short run, the value of this long/short portfolio is unaffected by small changes in the stock price.

The previous section explained why firms may find it cost-effective to issue underpriced convertible bonds. There is another issue that needs to be addressed as well: Why don't all fund managers (investors) purchase these underpriced securities? The answer lies in the very nature of convertible bonds, which are not stocks or bonds but a hybrid of the two. As such, many traditional money managers do not have a natural place in their portfolios for them. Furthermore, taking advantage of the mispricing of convertible securities requires managers to hedge a number of risks. This requires special skills, which traditional money managers may not possess. In addition, the investment strategy may not fall within the mandate of many traditional investment managers. It is useful to consider this arbitrage opportunity in terms of the earlier discussion regarding the limits to arbitrage. For example, this strategy typically relies on leverage, which could expose the portfolio to leverage risk. Also, since the underlying stock has to be shorted, the manager faces the risk that the shorted stock is called back, leaving the portfolio unhedged.

The hedging strategy is aimed at creating a market-neutral portfolio using long and short positions in the bond and the stock, respectively. This position is a hedge against either a rise or a fall in the stock price, but it still allows the investor to benefit from the long position in the underpriced convertible bond. In fact, available empirical evidence supports the notion that convertible arbitrage portfolios have low betas when there are small changes in the underlying stock price. Depending on the convertible arbitrage strategy pursued, the strategy may have a positive or negative beta to large changes in the underlying stock.

Convertible arbitrage not only entails buying convertible bonds but also requires a simultaneous short position in the underlying equity. The short position can provide several benefits. First, it hedges the equity risk embedded in a convertible bond—especially if the option is not deep-out-of-the-money. Second, since firms can time the issuance of convertible bonds, the bonds are likely to be issued when the firm considers its shares to be fairly valued or overvalued. Shorting an overvalued stock has the potential to be another source of short-term return to the arbitrageur, as well as interest earned on the short sale proceeds in the form of a short stock rebate. Finally, the short position in the equity provides a natural way of creating leverage. However, shorting opportunities may be limited and costly, and require a unique skill set. Only sophisticated relative value fund managers can take full advantage of this arbitrage opportunity.

Since convertible arbitrage is primarily a hedged strategy, it is necessary to continuously adjust the hedge ratio as relevant parameters such as stock price, return volatility, interest rates, and time to maturity change. This continuous adjustment of the portfolio puts an additional burden on the fund manager in terms of cost of trading and limits entry for typical bond investors.

27.2.4 A Convertible Bond Example

Convertible bonds give holders the right to exchange the bonds for common shares of the issuer at a stated ratio during a particular period. They are complex securities that blend the characteristics found in equity, debt, and options. As a result, traders and arbitrageurs have developed specialized terminology to describe various aspects of the marketplace. For the sake of simplicity, assume it is now January 1, 2016, and a hypothetical convertible bond is denoted as an XYZ convertible 2% coupon with a maturity date of December 31, 2020. Exhibit 27.3 describes the various parameters that characterize this convertible bond.

The fixed-income features of the convertible bond are:

- The *issuer* is the XYZ Company Inc., a company with a BBB *rating*.
- The convertible bond pays a 2% *annual coupon*, with the first coupon paid in exactly one year.
- There is no *accrued interest*, as the convertible bond has just been issued.
- The convertible bond has a five-year *time to maturity*.

EXHIBIT 27.3 Summary of the Terms Offered by the XYZ Convertible 2% 2020 Bond

Fixed-Income Features	
Issuer	XYZ Company Inc.
Rating	BBB
Coupon	$C = 2\%$ (annual)
Issue date	January 1, 2016 (today)
First coupon date	December 31, 2016 (in one year)
Accrued interest	0
Maturity	December 31, 2020 (in $T = 5$ years)
Nominal value	$1,000
Risk-free rate	$R_f = 4\%$ per year
Issuer credit spread	$CS = 400$ basis points (bps) above the risk-free rate
Equity Features	
Issuer	XYZ Company Inc.
Stock price	$S_0 = \$100$ per share
Stock price change volatility	$\sigma = 30\%$ per year
Stock dividend	None
Conversion Features	
Conversion ratio	$CR = 8$
Conversion price	$125 = Face value/conversion ratio = $1,000/8
Call protection	None
Market Valuation	
Convertible market price	90 (i.e., 90% of face value)
Parity	80 (i.e., 80% of face value) = (Conversion ratio × stock price)/convertible bond price
Conversion premium	12.50% = (90–80)/80

- The *nominal* or *par value* of each bond is \$1,000. This is the amount for which each bond can be redeemed at maturity.

The convertible bond can be converted into shares of the stock of the issuer. These shares have the following characteristics:

- The issuer is the XYZ Company Inc., the same issuer as for the convertible bond.
- The stock price is currently \$100 per share.
- The historical volatility of the stock price is 30% per year.
- The stock pays no dividend.

The terms of the conversion are fixed in the convertible bond's indenture as follows:

- The conversion ratio denotes the number of shares obtained if one converts \$1,000 of the face value of the bond. In this example, each bond with a \$1,000 face value can be converted into eight ordinary shares. The conversion ratio is therefore 8. This number usually remains fixed through the life of the instrument unless stock splits, special dividends, or other dilutive events occur.
- The conversion price denotes the price at which shares are indirectly purchased via the convertible security. This price is equal to the par value of the convertible security divided by the conversion ratio, or \$1,000/8 = \$125.
- **Call protections** grant the issuer the right to call back the convertible bond before its stated maturity. This can be either a **hard call**, in which the issuer can call the bond at a prefixed price regardless of any other circumstances, or a **soft call**, in which the issuer can call the bond only if the equity price has risen significantly above the strike price or some other hurdle rate. In this example, for the sake of simplicity, it is assumed that there is no call protection.

The valuation parameters of the convertible bond that are observable are:

- The convertible price denotes the quoted price of the convertible bond, which is usually expressed as a percentage of the nominal value (consistent with the traditional bond market). In this example, it is assumed that the convertible is quoted at 90% of its face value.
- **Parity** is the total value of the shares into which the bond can be converted based on the current market price of the shares. In this example, it is calculated as 8 shares per bond × \$100 per share = \$800. Parity is normally quoted as a percentage of the par amount of the bond; thus, \$800/\$1,000 = 80%, or simply 80. Note that a convertible bond is said to be **in-the-money** when its parity is higher than its face value.
- The **conversion premium** is the difference between the convertible bond price and parity, expressed as a percentage of parity. In this example, the conversion premium is (90 − 80)/80 = 12.5%. The premium expresses how much more an investor has to pay to control the same number of shares via a convertible. This premium also gives an indication of how a convertible should perform in relation to the underlying shares. All else being equal, convertibles with very low

premiums will be much more sensitive to movements in the underlying share price (i.e., parity) than convertibles with higher premiums.

In addition to these standard features, convertible bonds may have more complex characteristics, some of which are described here:

- Zero-coupon convertible bonds are issued at a deep discount to par value and are redeemable at par. The most famous examples of such bonds are liquid yield option notes (LYONs), which are both callable (redeemable prior to maturity by the issuer) and putable (redeemable prior to maturity by investors).
- Mandatory conversion securities (MCSs) are convertibles whose conversion is mandatory at some stage. They tend to trade and behave like shares, although some may have additional features; for example, a preferred equity redemption cumulative stock (PERCS) is a mandatory preferred convertible with a preset cap level, above which the conversion ratio is adjusted to keep the total return payoff constant (i.e., as the underlying stock price rises, the PERCS becomes convertible into fewer and fewer underlying shares).
- Convertible preferred shares are preferred stocks that include an option for their holder to convert them into a fixed number of common shares, usually following a predetermined time span.

In the following valuation exercise, these complex features are largely ignored, with the focus on the arbitrage of plain-vanilla convertible bonds.

27.2.5 Valuation of Convertible Securities: Component Approach

To identify convertible bond arbitrage opportunities, one must be able to calculate the fair value of a convertible bond. The component approach is the most intuitive valuation approach for simple convertible bonds and was emphasized in Chapter 19 of the CAIA Level I book. It essentially divides the convertible bond into a straight bond component and a call option:

$$\text{Convertible Bond} = \text{Straight Bond} + \text{Call Option on the Underlying Stock} \quad (27.1)$$

The straight bond component refers to the pure fixed-income portion of the convertible bond. It ignores the conversion possibility, and its value is easily obtained by discounting all future expected cash flows (coupons and final repayment) at an appropriate discount rate (the risk-free rate plus a credit spread). In the following example, the discount rate of 8% is composed of a 4% risk-free rate added to the credit spread of 4%. Exhibit 27.4 illustrates the computation using a discount rate of 8% to calculate the total present value of the XYZ convertible bond as $760.44. Note that the coupon rate of interest, 2%, is far below the discount rate of 8%. Issuing convertible bonds rather than straight debt reduces the current-year interest expense of the firm, as investors are willing to earn a lower coupon rate of interest in exchange for the value provided through the embedded equity call option.

EXHIBIT 27.4 Expected Cash Flow (CF) Decomposition for a Convertible Bond

Time (Years)	0	1	2	3	4	5
Cash flows		$20	$20	$20	$20	$1,020
PV(CF) @ 8%		$18.52	$17.15	$15.88	$14.70	$694.19
Total PV	$760.44					

This pure bond price is, in a sense, the minimum value of a convertible bond. It is unaffected by the stock price level unless the latter falls so much that the issuer's ability to pay its debt obligations is called into question.

The option component considers the conversion features of the convertible bond only. It is essentially an option to buy a certain quantity (the conversion ratio) of shares of stock by paying the value of the convertible bond. Prices of regular options are frequently calculated using the Black-Scholes formula. Using the Black-Scholes model, the theoretical price for the option component of the XYZ bond is $202.24, which is an options price of approximately $25.28 for each of the eight shares into which the bond is convertible. The theoretical price of the convertible bond is obtained by summing the theoretical prices of its components: $760.44 + $202.24 = $962.68, or 96.27% of its face value. Note that the market value of the bond was reported to be 90% of its face value.

The major drawbacks of the outlined component approach are threefold: (1) It uses the Black-Scholes model, which is valid only for European options and cannot deal with early termination clauses; (2) it does not take the credit risk of the issuer into consideration; and (3) it does not account for special conditions and contractual covenants (e.g., callability).

27.2.6 Valuation of Convertible Securities: Binomial Model

An alternative way of determining the price of a convertible bond is via a binomial tree approach, introduced by Cox, Ross, and Rubinstein (1979) and discussed in CAIA Level I. This approach is widely used in practice because it can deal with a wide range of contractual specifications while still remaining relatively simple.

Let us assume that the stock price process follows a binomial random walk. That is, over the next period, the stock price can either go up by a multiplicative factor u or go down by a multiplicative factor d, where $u > d$ (see Exhibit 27.5).

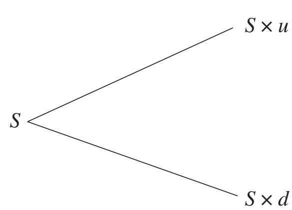


EXHIBIT 27.5 Possible Stock Price Movements (One Period)

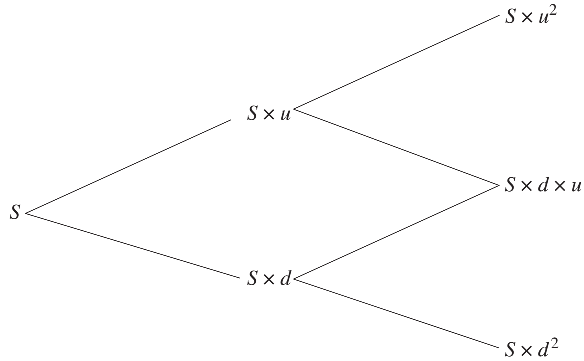


EXHIBIT 27.6 Possible Stock Price Movements (Two Periods)

This process can be repeated as many times as needed and is made manageable by its recombining nature, meaning that if the price moves up and then down, the price moves back to its prior level. For instance, over two periods, the tree in Exhibit 27.6 is obtained, representing the evolution of discrete possible underlying stock prices into three ending prices.

The parameters u and d are proportional to the volatility of the underlying asset and must be specified. As suggested by Cox, Ross, and Rubinstein (1979), set $u = \exp(\sigma\sqrt{\Delta t})$, where Δt is the length of one period in the tree, σ is the volatility of the stock returns, and $d = 1/u$. This ensures that the resulting expected volatility of percentage changes in the stock price is σ.

As discussed in Chapters 6 and 10 of the CAIA Level I book, binomial models can be used to price derivative products. Here, the procedure for calculating derivative prices using the binomial tree approach is briefly discussed. The same approach can be used to price a convertible bond.

The first step in calculating an arbitrage-free price for a derivative security is to calculate the payoff of the derivative based on the binomial tree for the underlying asset. Suppose there is a one-period call option on the stock that is depicted in Exhibit 27.5. As a result, the payoff from this call option will be $\max(S \times u - K, 0)$ in the top node and $\max(S \times d - K, 0)$ in the bottom node, where K is the strike price of the call option. The next step is to calculate the arbitrage-free price of the call option using the expected payoff of the option under risk-neutral probability discounted at the riskless rate.

As discussed in Chapters 6, 10, and 26 of the CAIA Level I book, risk-neutral probabilities are used to find prices of securities as if investors were risk neutral. It is important to note that risk-neutral probabilities differ from the actual probabilities of up and down movements. They are constructed for the sole purpose of estimating the arbitrage-free price of a derivative instrument. This means that, given the assumed values for u and d, the resulting price would not provide an arbitrage opportunity to market participants. The calculated option prices are consistent with actual market prices of the underlying assets because, to begin with, these prices are used to calculate risk-neutral probabilities.

More specifically, it is known that the current price of the stock is its expected payoff under the risk-neutral probability discounted at the riskless rate. That is,

$$S = \frac{[p \times S \times u + (1 - p) \times S \times d]}{r} \tag{27.2}$$

where p is the risk-neutral probability of the up movement, and r is 1 plus the riskless rate. Equation 27.2 can be solved for the risk-neutral probability:

$$p = \frac{r - d}{u - d} \tag{27.3}$$

Clearly, to prevent arbitrage, we must have $d < r < u$.

Example: Suppose $r = 1.05$, $u = 1.1$, and $d = 1/u$. The risk-neutral probability of an up movement is 0.738; therefore, the probability of a down movement is 0.262.

These risk-neutral probabilities can be used to calculate the price of the call option.

Example: Suppose the stock price is 100 and the call option's strike price is 100 as well. The arbitrage-free price of the option will be:

$$C = \frac{0.738 \times \max(100 \times 1.1 - 100, 0) + 0.262 \times \max(100/1.1 - 100, 0)}{1.05} \tag{27.4}$$

Having discussed the basic approach to the pricing of derivative securities using the binomial approach, we return to the XYZ convertible bond. We will draw the binomial trees for the underlying stock as well as the payoff from the convertible bond.[5] To analyze the convertible bond, the life of the tree is set equal to the life of the convertible bond (five years). For the sake of simplicity, we consider a tree made of $N = 5$ periods of $t = 1$ year, which implies that the convertible bond pays its coupon each period. The size of the up move is $u = \exp(0.30) = 1.3499$, while the down move is $d = 1/u = 0.7408$. The resulting stock price tree is given in Exhibit 27.7.

$t = 0$	$t = 1Y$	$t = 2Y$	$t = 3Y$	$t = 4Y$	$t = 5Y$
					448.17
				332.01	
			245.96		245.96
		182.21		182.21	
	134.99		134.99		134.99
100.00		100.00		100.00	
	74.08		74.08		74.08
		54.88		54.88	
			40.66		40.66
				30.12	
					22.31

EXHIBIT 27.7 Binomial Tree for the Stock Price

$t = 0$	$t = 1Y$	$t = 2Y$	$t = 3Y$	$t = 4Y$	$t = 5Y$
					358.54
				265.61	
			196.77		196.77
		145.77		145.77	
	107.99		107.99		107.99
80.00		80.00		80.00	
	59.27		59.27		59.27
		43.90		43.90	
			32.53		32.53
				24.10	
					17.85

EXHIBIT 27.8 Binomial Tree for Convertible Bond's Parity

As discussed, the binomial tree structure can be used to extract the risk-neutral probability of the stock price moving up (increasing), p, or going down (decreasing), $(1 - p)$. The risk-neutral probability for the problem at hand is computed as follows:

$$p = (r - d)/(u - d) = (1.04 - 0.7408)/(1.3499 - 0.7408) = 0.491 \quad (27.5)$$

To build a tree for the convertible bond, note that the conversion value of the convertible bond is given by the stock price times the conversion ratio; for example, the conversion value of the bond in the first node is 800, or 80% of face value. Thus, the associated parity tree can be easily constructed, as shown in Exhibit 27.8. For instance, in the uppermost node of the last period, the conversion value of the bond is 358.54 percentage of face value ($8 \times 448.17/1{,}000$).

At time $t = 5$ years, the pure bond price of the convertible bond is its repayment value ($1,000) plus the last coupon ($20). Conversion into shares should occur only if the conversion value is higher than $1,020. It is easy to verify that this is the case when the final stock price is $448.17, $245.96, or $134.99 (i.e., for the top three end nodes in Exhibit 27.7, when parity is greater than 100). For the bottom three end nodes, no conversion should occur. Therefore, at time $t = 5$, there is a 100% probability of converting for the top three nodes, and a 0% probability of converting for the bottom three nodes. Given the tree that provides the payoff for the convertible bond and the risk-neutral probabilities, the arbitrage-free price of the bond can be calculated.

27.2.7 Convertible Bond Behavior at Various Stock Price Levels

The overall profile of a convertible bond can be obtained by calculating the theoretical prices of its components for various levels of the stock price.

As illustrated in Exhibit 27.2, there are four possible states for a convertible bond:

1. Distressed or junk convertible: **Distressed** or **junk convertibles** occur when the stock price is so low that it indicates substantial doubt about the issuer's

ability to meet its convertible debt obligations. The call option is likely worth zero, and the convertible trades like a distressed bond would. The parity is typically between 0% and 30% of the face value. A distressed convertible is said to be deep-out-of-the-money.

2. Bondlike or busted convertible: This occurs when the stock price is low enough that conversion is unlikely. The value of the call option is negligible, and the convertible bond trades essentially similar to a straight bond with no equity sensitivity. The parity is typically between 30% and 80% of the face value. A busted convertible is said to be out-of-the-money.
3. Hybrid convertible: This takes place when the stock price is high enough that the option to convert gains value. The parity is typically between 80% and 120% of the face value, and the hybrid convertible is said to be at-the-money.
4. Equity-like convertible: This occurs when the stock price is extremely high, making conversion likely. The convertible trades more like a stock than a bond. Its equity sensitivity is high, whereas its fixed-income sensitivity is low. The parity is typically above 120% of the face value, and the convertible is said to be deep-in-the-money.

It may be recalled that the price of a call option can be broken into two parts: the intrinsic value and the time value (i.e., the optionality value). The optionality value tends to be highest for at-the-money options. For this reason, convertible arbitrage managers generally prefer hybrid convertibles that exhibit both fixed-income and equity sensitivities. Hybrid convertibles tend to have greater profit potential because the optionality value of a convertible bond is highest when the embedded option is at-the-money.

27.2.8 Greeks of Convertible Bonds

The Greeks are measures that represent the sensitivities of convertible bonds in relation to a variety of factors. The two critical measures of sensitivities as far as convertible arbitrage is concerned are delta and gamma.

Delta measures the sensitivity of the value of a derivative security (e.g., a convertible bond) to changes in its underlying asset (i.e., its stock price or parity level). Mathematically, it is the first-order partial derivative of the convertible bond value with respect to the underlying stock price. (Note that the following expression is strictly correct if the conversion ratio is 1. More precisely, delta measures the sensitivity of the bond to changes in its parity value.)

$$\text{Delta} = \delta = \frac{\partial\,\text{Convertible Value}}{\partial\,\text{Stock Price}} \tag{27.6}$$

Since the underlying asset of a convertible bond is the stock price times the conversion ratio, the delta of a convertible is calculated using the sensitivity to parity. Delta corresponds to the slope of the convertible bond price curve displayed in Exhibit 27.2. A steeper slope indicates a higher sensitivity to the underlying stock price. When a convertible bond is deep-in-the-money, it begins to trade like the stock, moving almost dollar for dollar with the stock price (equity-like, delta = 1). Meanwhile, out-of-the-money convertibles do not move much in absolute dollar terms (bondlike, delta = 0) unless there are some serious bankruptcy concerns (distressed).

So far, it has been implicitly assumed that the volatility of the underlying stock and the risk-free interest rate are constant. In reality, changes to the time to maturity, underlying stock return volatility, or risk-free interest rate can affect the delta of the convertible bond, even if the price of the underlying stock remains the same. In fact, the sensitivity of delta to changes in any of these parameters depends on whether and to what extent the embedded option is in- or out-of-the-money.

Delta is a critical variable for convertible arbitrageurs, for it indicates how many shares of stock one should sell short to hedge the equity risk of a long position in a convertible bond. Delta provides a linear approximation for the change in the convertible bond value. It is relatively accurate for small variations of the underlying stock price, but it fails to capture the nonlinearity of the convertible bond profile. To correct for this, it is necessary to introduce a convexity adjustment.

Gamma measures the rate of change of the delta as the stock price changes. Mathematically, it is the second derivative of the convertible bond value with respect to the underlying stock price (assuming a conversion ratio of 1):

$$\text{Gamma} = \Gamma = \frac{\partial^2 \text{ Convertible Value}}{\partial \text{Stock Price}^2} = \frac{\partial \text{Delta}}{\partial \text{Stock Price}} \tag{27.7}$$

Gamma is important for convertible arbitrageurs, as it indicates how much delta changes with regard to changes in the stock price. If gamma is large (i.e., delta is very sensitive to the underlying price), the portfolio will need to be adjusted frequently to maintain delta neutrality. If gamma is small (i.e., delta changes slowly), the portfolio will need to be adjusted less frequently to keep the portfolio delta neutral.

Note that gamma changes as delta changes. Gamma is normally larger for at-the-money convertibles and becomes progressively smaller for both in- and out-of-the-money convertibles. In the case of distressed convertibles, gamma can even become negative.

In addition, gamma also changes as time passes and volatility fluctuates, but in a more complex way than delta; time passing or volatility decreasing normally increases the gamma of at-the-money convertibles but decreases the gamma of in-the-money and out-of-the-money convertibles.

27.2.9 Convertible Bond Arbitrage Strategies: Implementation

We have seen that the first step in implementing a convertible bond arbitrage strategy is to identify underpriced convertibles. The component approach and the binomial approach are essential tools in implementing the first step. In implementing a convertible arbitrage strategy, managers consider the following preferences:

- Underlying stock preferences: Managers prefer stocks with high return volatility and a low dividend yield, and that are liquid and easy to borrow.
- Convertible bond preferences: Managers generally prefer convertible bonds, or converts, that have low premiums relative to their conversion value, credit risk that can be managed, and good covenant protection.

Depending on the type of convertible bond being employed, the strategy will have different risk-return profiles.

- Deep-in-the-money converts: A convertible bond strategy involving such bonds (equity-like in Exhibit 27.2) is similar to a synthetic put. A **synthetic put** is a portfolio that has the same payouts as a put but without a pure put as a component. Implementing a convertible bond arbitrage strategy with a deep-in-the-money convert is also called a **cash-flow strategy**, which generates the bulk of its return from the steady cash flows generated by the position. In this strategy, the hedge ratio is high and the goal is to create a cheap synthetic put and collect cash flows in the process. Recall from put-call parity that a long position in a call plus a short position in a stock and a long position in a bond create a synthetic put. Therefore, if the option embedded in the convertible bond is underpriced, this strategy will allow the manager to create a synthetic put at a relatively low cost. This strategy will generate a steady cash flow as the investor earns the coupon rate of the bond and pays the dividend yield of the stock. Also, similar to a put, it will benefit from a sharp drop in stock prices or an increase in volatility. Similar to other options, this strategy will lose some money due to time decay.
- At-the-money converts: A strategy involving at-the-money convertible bonds (hybrid in Exhibit 27.2) benefits if the implied volatility of the embedded option is below the expected volatility of the stock price. Strategies based on identifying mispriced securities resulting from divergences between expected and implied volatilities are referred to as **volatility trading**. Therefore, it is essential that the manager be able to identify whether the embedded option of the convertible is indeed underpriced. Although the position will be generating a cash flow going forward, the potential decline in the option premium may offset this gain if there is a decline in the volatility of the underlying asset.
- Distressed converts: A convertible bond strategy involving such bonds (distressed in Exhibit 27.2) benefits from identifying converts that are underpriced because of their high probability of default. Managers who specialize in this segment of the market must have the skills to manage credit risk. In this strategy, the hedge ratio may be high not because the equity risk of the convert is being hedged but because the short position in equity is being used to hedge the credit risk of the long convert position. The value of the embedded option of the convertible is small in this case, and the bond trades similarly to a plain distressed debt. The potential source of return comes from the fact that traditional distressed managers may not be interested in investing in hybrid securities, and therefore the issue could be underpriced. Of course, the investor will experience significant

benefits if the firm recovers, as the bond floor and the embedded option will experience a substantial rise in value.

27.2.10 Convertible Bond Arbitrage Strategies: Managing Four Non-Equity Risks

So far, our focus has been on the equity risk of convertible bonds and models that can help us identify the appropriate hedge ratios. However, successful implementation of this strategy requires the manager to pay close attention to other risk factors as well, such as these four:

1. Credit risk: As previously mentioned, a majority of convertible bond issues are below investment grade. This means that default risk can play an important role in the performance of this strategy. A short position in the underlying stock can reduce the portfolio's exposure to credit risk. Managers may choose to use other tools, such as credit default swaps, to have more control over the credit risk exposure of their portfolios.
2. Interest rate risk: This risk is particularly important for long-term convertibles that are not deep-in-the-money. The short position in the underlying stock may not provide any protection against this risk; therefore interest rate derivatives may be used to hedge this risk.
3. Event risk: Certain events—such as mergers, an unexpected increase in dividends, a short squeeze on the borrowed stock, or a call of the bonds—could have a significant negative impact on the performance of this strategy. For example, if the issuing company is subject to a takeover by a larger firm, the post-merger volatility of the underlying stock could become much lower, leading to a decline in the value of the embedded option. An increase in the dividend yield of the stock will make the short position more costly, negatively affecting the performance of the strategy. Also, if the convertible bond is callable, there is a risk that a convertible bond that is selling at a premium will be called at the par value by the issuer.
4. Crowding effect risk: Hedge funds constitute the largest group of investors in convertible bonds. Since convertible bond markets are not highly liquid, there could be market disruptions if a large number of hedge funds suddenly attempt to liquidate their positions. This is indeed what happened in 2005 and in 2008, when a large number of leveraged convertible arbitrage hedge funds faced redemptions and had to sell portions of their convertible bond portfolios. The resulting market disruptions led to significant losses for this strategy.

27.2.11 Five Sources of Convertible Arbitrage Returns

A convertible arbitrage strategy generates returns from a variety of sources, with their relative importance changing depending on the nature of the strategy. The following five sources are present in most convertible arbitrage strategies:[6]

1. Coupon and face value payments: These represent the income and return of principal from holding a long position in the bond.

2. Dividend payments: These represent the costs associated with shorting the stock. The manager must make the dividend payments to shareholders from whom the stock has been borrowed.
3. Rebates on short interest: These represent the income paid by a stock lender to a borrower, which is interest on the borrower's cash collateral in excess of the stock loan fee.
4. Arbitrage returns: These represent the returns due to the underpricing of the convertible's embedded option, which include profits from rebalancing the short stock position as delta changes.
5. Leverage cost: Since the profit margin on most convertible arbitrage trades is rather small, many managers use leverage to increase their returns. As long as investors can tolerate the increased risk and the cost of leverage is less than the rate of return from the unlevered strategy, increased leverage will enhance the fund's return.

APPLICATION 27.2.11

Exhibit 27.9 presents the five sources of convertible arbitrage returns using a numerical example. In this example, the initial price of a 5% coupon convertible bond is assumed to be 108% of face value. The total amount of capital invested in this trade is $1,080,000, of which $202,500 is equity and $877,500 is borrowed. This means that 1,000 bonds with a total face value of $1,000,000 are purchased. The bond's conversion ratio is 34.783 shares of stock. The delta of the option is estimated to be 0.75. The interest rate on short proceeds is 1.25%. Finally, the current price of the common stock is $26.625. It is assumed that the manager closes these positions after one year by selling the bonds out $120 and buying back the stock at $31.00. The five sources of return are shown in Exhibit 27.9.

EXHIBIT 27.9 Sources of Returns for Convertible Arbitrage

Source of Return/Cost	Return/Cost	Assumptions/Computations
1. Interest income from long bond	$ 50,000	5% coupon on $1,000,000 face value
2. Short rebate from short stock		Shares shorted = 34,783 × 0.75 = 26,000 (approximate)
		Total amount shorted = $26,000 × $26.625 = $692,250
	$ 8,653	Rebate = $692,250 × 1.25% = $8,653
Cost of leverage		2% cost of borrowing
	$ (17,550)	$877,500 × 2% = $17,550
Dividend payments on		1% dividend yield on shares
shorted shares	$ (6,922)	$692,250 × 1% = $6,922

(continued)

EXHIBIT 27.9 (*Continued*)

Source of Return/Cost	Return/Cost	Assumptions/Computations
Total annual cash flows	$ 34,181	
Bond return	$ 120,000	Bonds purchased at 108 and sold at 120
Stock return	$ (113,750)	Stock shorted at $26.625 and covered at $31.00
Total arbitrage cash flow	$ 6,250	
Total cash flow	$ 40,431	
Bond income	4.63%	$50,000 income on $1,080,000 investment
Short interest rebate	0.80%	$8,653 income on $1,080,000 investment
Dividend payment	–0.64%	$6,922 cost on $1,080,000 investment
Arbitrage return	0.58%	$6,250 income on $1,080,000 investment
Cost of leverage	–1.63%	$17,550 cost on $1,080,000 investment
Return on assets	3.74%	$40,431 income on $1,080,000 investment
Return on equity	19.97%	$40,431 income on $202,500 equity investment
Contribution of leverage	16.22%	19.97% – 3.74%

Source: AIMA (2006).

27.2.12 Market Size and Historical Performance

Historically, convertible arbitrage started as a niche business for dedicated proprietary trading desks in large investment banks. Convertible arbitrageurs typically bought cheap convertible bonds and hedged their market risk by selling short the underlying stocks. Subsequently, thanks to the development of sophisticated option pricing models and the availability of credit derivatives, the strategy expanded to include volatility and credit trading elements.

Hedge Fund Research (HFR) reports that as of the third quarter of 2015, dedicated convertible arbitrage funds represented approximately $50.6 billion—or 1.76% of total assets, down from 3% in 2011. Among relative value funds, convertible arbitrage made up 6.5% of assets. Though quite small in comparison to equity hedge funds—including long/short funds, which made up more than 28% of hedge fund assets—convertible arbitrage shares important features with a variety of hedge fund strategies and serves as a valuable example. Exhibit 27.10 displays the performance of the CISDM Convertible Arbitrage Index as well as that of global equity and fixed-income indices for 2005–15.

27.3 PAIRS TRADING AND MARKET NEUTRALITY

This section considers another form of relative value arbitrage: pairs trading, or statistical arbitrage. Unlike convertible arbitrage, in which an arbitrageur takes a

EXHIBIT 27.10 Risk-Return of Convertible Bond Arbitrage Strategy

2005–15 (USD)	Annualized Mean Return	Annualized Std. Dev.	Maximum Drawdown	Information Ratio	Correlation with MSCI World	Correlation with Barclays U.S. Agg	Correlation with Barclays U.S. High Yield
CISDM Convertible Arbitrage	5.46%	6%	–22%	0.87	0.62	0.25	0.82
MSCI World	6.30%	16%	–56%	0.39	1.00	0.08	0.75
Barclays Global Aggregate	3.08%	6%	–10%	0.56	0.41	0.70	0.36
Barclays Global High Yield	7.34%	11%	–33%	0.68	0.80	0.27	0.98
Barclays U.S. Aggregate Bond	4.30%	3%	–4%	1.34	0.08	1.00	0.24
Barclays U.S. Corporate High Yield	6.91%	10%	–33%	0.68	0.75	0.24	1.00

Source: Bloomberg and authors' calculations.

EXHIBIT 27.11 Randomly Generated Numbers

Observations	1	2	3	4	5	6	7
X	0.74422	0.37724	0.09228	0.87045	0.32695	0.62064	0.18186
Y	0.46229	0.65782	0.00749	0.44523	0.71240	0.49302	0.87969

simultaneous long and short position in two securities—a bond and a stock—from the same issuer, in pairs trading, the arbitrageur invests long and short in pairs of related securities of different issuers. Similar to the convertible arbitrage strategy, in which long and short positions are used to create a return stream that is not highly exposed to changes in equity markets, returns from pairs trading strategies are expected to have low correlations to equity markets. The following paragraphs first discuss the broad concept of an equity market-neutral strategy. Then the discussion focuses on pairs trading as a specific form of an equity market-neutral strategy, explaining the concept of co-integration and contrasting it with correlation. Finally, we discuss the risk aspects of this strategy.

27.3.1 Equity Market-Neutral Strategy: General Framework

Simply put, an equity market-neutral strategy refers to a set of strategies whose returns are independent from equity market fluctuations. The correlation coefficient is the most common measure of independence used by the industry to determine if a strategy is market neutral. It is important to note that when two return streams are uncorrelated, it does not mean that they are independent. Basically, correlation is useful in determining whether the two return streams are linearly related to each other. To see this, consider the two random variables that appear in Exhibit 27.11.

It can be verified that the correlation between X and Y is zero. However, these two random variables are not independent. If they were independent, then the correlation between X and $f(Y)$, where $f(\bullet)$ is some arbitrary but well-defined function, would be zero. In particular:

$Correl(X, Y) = 0$	Linearly independent
$Correl\,(X, f(Y)) = 0$	Independent if holds for any f

In the preceding example, it can be verified that the correlation between X and Y^2 is 0.3. This means that these two random variables are not independent.

This may appear to be too technical and not relevant to the study of an equity market-neutral strategy. However, by recognizing that returns on some derivative securities are nonlinear functions of returns on their underlying assets or that through dynamic trading strategies one can replicate returns on derivative securities, it can be seen that a strategy whose return is uncorrelated with the market may not be independent of the market. For example, for small changes, the correlation between a fund's returns and the market could be almost zero; but for large changes, the correlation could be large and statistically significant. This means that one has to examine the market neutrality of a fund's return under a variety of market conditions in order to have a high degree of confidence that the fund is truly market neutral.

It is important to emphasize that even if a portfolio is uncorrelated with or independent from the overall market, it might have significant exposure to specific industries. In addition, it is likely to have exposure to other sources of risk in the economy. For example, it might be exposed to credit, currency, or volatility risk. Therefore, an equity market-neutral portfolio does not mean a riskless portfolio, and the appropriate rate on such a portfolio will not be the London Interbank Offered Rate (LIBOR) or other proxies for the riskless rate. The capital asset pricing model (CAPM) is often used to argue that the appropriate rate of return on a portfolio with zero equity beta is the riskless rate. However, the CAPM assumes that the market portfolio is the only source of risk. In a more general setting, in which there are many sources of risk through time, zero beta with respect to the market portfolio does not mean the portfolio has no systematic risk.

The hedge fund industry has developed a few specific terms around the concept of the equity market-neutral strategy:

- Monetary neutrality: A portfolio that is **monetary neutral** has equal long and short exposures to a specified currency. For example, the portfolio has equal amounts of euro exposure in long and short positions. A euro-neutral manager expects the long positions to outperform the short positions. The strategy is not expected to generate returns that are uncorrelated with the overall market, but it is expected to have a low correlation with the market. A euro- or dollar-neutral strategy is approximately self-financing, as the proceeds from the short positions can, in principle, be used to fund the long positions. This means the firm's equity can be invested in cash or cash-equivalent securities. The strategy may not be fully self-financing, as not all the proceeds from the short sales may be available to the portfolio manager.
- Beta neutrality: A **beta neutral** portfolio generates returns that are uncorrelated with the market risk associated with the specified beta. In this strategy, the equity market betas of long and short positions are equal to each other. Note that this does not mean that equal amounts are invested in long and short positions. For instance, if a long position consists of low-beta (e.g., value) stocks, and a short position consists of high-beta (e.g., growth) stocks, then the size of the long position needs to be levered up in order to match the high beta of the short position.
- Sector neutrality: A **sector neutral** portfolio generates returns that are uncorrelated with economic sectors, including industries or industry groups. For example, a fund manager may take long positions in the stocks of energy firms and short positions in the stocks of technology firms. Even if the two positions are of the same size and have the same beta with respect to the changes in the overall market, the portfolio has significant positive and negative exposures to the energy and technology sectors. On the other hand, a manager may choose to create long and short positions using energy stocks. This portfolio will have little exposure to changes in the overall energy sector but may not be beta neutral.

An important question that arises is whether an equity market-neutral strategy will generate any return. After all, if markets are efficient, then a truly market-neutral strategy should generate zero return. An equity market-neutral portfolio may generate positive returns in an efficient market, but one could argue that those returns are compensations for exposures to systematic risks other than equity risk (e.g., credit, currency, volatility, or liquidity risk). Alternatively, one can argue that markets are not always efficient, and pockets of inefficiency arise in financial markets. The goal of equity market-neutral strategies is to exploit these inefficiencies while not exposing the portfolios to substantial amounts of other systematic risks.

27.3.2 Pairs Trading: Conceptual Framework with Four Steps

The development of the pairs trading strategy is attributed to Nunzio Tartaglia, who, along with a team of statisticians and mathematicians, worked at Morgan Stanley in the 1980s. Their objective was to use quantitative models to identify and implement arbitrage opportunities in a systematic and algorithmic manner. One of the strategies successfully developed by this team was statistical pairs trading.[7]

The objective of a pairs trading strategy is to identify securities that have moved together in the past, and to create pairs of these securities with the belief that they will continue to move together in the future (e.g., stock prices of Exxon Mobil and Chevron, two large oil companies). The relative mispricing of the two securities is represented by the notion of the spread between the two returns; hence, the greater the spread, the greater the size of the mispricing and the greater the profit potential. A portfolio consisting of long and short positions in the two securities is constructed. Once the prices of the two stocks converge back to the expected relationship, the manager exits the positions. If the underlying assumption of the strategy, which is that the two prices should converge and that the spread should be zero, is valid, then the resulting portfolio return will typically have low correlation with changes in the overall market; that is, it may not be beta neutral, but the beta will typically be low. Also, since the two stocks are likely to come from the same industry, the portfolio is likely to have low exposure to the sector. In most instances, equal amounts are invested in long and short positions; therefore, the strategy tends to be monetary neutral.

The idea of selling (shorting) overvalued stocks and buying undervalued stocks on a consistent basis is too general and too ambitious to work for markets that are relatively efficient. Even in highly inefficient markets, identifying undervalued and overvalued securities may not necessarily lead to a profitable strategy. If the market remains inefficient, there is no reason for prices to converge to their fundamental values. The pairs trading strategy is based on the implicit assumption that market prices may deviate from the fundamental values but will converge to those values in a relatively short period of time.

The implementation of a pairs trading strategy consists of the following four steps:[8]

1. Identify the candidate pairs: In this step, statistical and fundamental analysis can be used to identify stock pairs that are closely related, such that the expected spread between the two returns is expected to be zero. The following paragraphs discuss one such statistical method: co-integration.
2. Identify pairs with divergent spread: Once candidate pairs are identified, their price paths are monitored to determine if the observed spread has exceeded a predetermined threshold.

3. Construct portfolio: Given the information collected in the previous two steps, the portfolio manager will have the estimates needed to create a long/short portfolio.
4. Exit strategy: This is perhaps the most difficult part of the strategy. Imagine that the previous steps have been taken and the portfolio has been constructed. The portfolio manager then observes that the spread between the two prices is actually increasing. There are two possible interpretations. First, the portfolio manager may consider this to be an even better opportunity and therefore adds to the positions. Alternatively, the portfolio manager may conclude that the underlying thesis of the trading has proven to be incorrect and therefore closes the positions. There is no systematic and precise way of knowing which of these two scenarios is correct.

27.3.3 Implementation of Pairs Trading

As mentioned in the previous section, the first step is to identify candidate pairs. One popular approach to finding candidate pairs is co-integration. The **co-integration approach** is a statistical technique that indicates the relationship between nonstationary time-series variables. Co-integration is used by traders to specify how the prices of securities may be related through time. It is important to point out that co-integration is different from correlation. Stocks that are highly correlated may not necessarily be good candidates for pairs trading. The following discussion will help clarify the differences between the two.

Consider two stocks with prices given as p_t and s_t. These stock prices are said to be co-integrated if a linear combination of the two is a stationary process. Loosely speaking, a process is said to be stationary if its statistical properties do not change through time; for example, its means and standard deviations remain the same. Therefore, the two stock prices are said to be co-integrated if the random variable u_t is stationary:

$$\ln(p_t) - a \times \ln(s_t) = u_t \tag{27.8}$$

Notice that the natural logs of prices rather than the prices themselves are used. The reason is that we are interested in the rate of return from the investment strategy, which is related to changes in the log of prices.

To see the idea behind Equation 27.8, suppose the logs of the two price processes are given by:

$$\ln(p_t) = \eta_{1t} + \varepsilon_t \tag{27.9}$$

$$\ln(s_t) = \eta_{2t} + v_t \tag{27.10}$$

Here, η_{1t} and η_{2t} are the common trends present in the two price processes, and ε_t and v_t are stationary random noises. If the common trends are related according to $\eta_{1t} = a \times \eta_{2t}$, then one can see that the linear combination given in Equation 27.8 will be stationary as long as a is a constant. There are well-known tests that allow one to determine whether two price series are co-integrated.

Depending on the correlation between the two random noises of Equations 27.9 and 27.10, the two price series could have different levels of correlations. However, regardless of the level of correlation, the logs of the two prices will be co-integrated if the nonstationary parts of the price series are represented by two related common trends.

There are other statistical methods for identifying potential candidates, and most traders try to supplement their statistical results with fundamental analysis. This ensures that the statistical results are not obtained by chance, and that there are real economic reasons for the two stock prices to move together.

Once the potential candidates have been identified, the portfolio manager will monitor their price movements through time. In addition, the portfolio manager will set predetermined thresholds, which will signal if the spread between two candidate stocks is wide enough to be exploited. The most common approach is to set the threshold proportional to the historical standard deviation of the spread. This means that if the spread has been historically very volatile, then the two prices have to diverge by a relatively large amount before a trade is triggered.

Finally, the portfolio manager must develop a precise exit strategy. For instance, the strategy would require the manager to exit the positions if the prices do not converge within a specified period or if the spread increases beyond a set level. Alternatively, the strategy could require the manager to exit the positions after a prespecified period of time no matter what has happened to prices.

One of the most cited research papers in this area is by Gatev, Goetzmann, and Rouwenhorst (2006), who find that annualized returns of up to 11% above Treasury rates can be generated by systematic pair trading strategies. The abnormal returns reported by this study remain significant even after accounting for transaction costs. The average holding period for the strategy was four months. The authors highlight the importance of diversification and recommend that in order to have relatively stable returns, a portfolio consisting of about 20 pairs should be constructed.

Another important study examines the sources of returns behind potential profits in pairs trading and how idiosyncratic news, common news, and liquidity affect the strategy's performance. In this study, Engelberg, Gao, and Jagannathan (2009) argue that idiosyncratic news, which temporarily decreases the liquidity in one of the stocks in a pair, increases the probability of finding profitable opportunities. Further, they explain that common news may lead to a faster reaction in one of the stocks, depending on the difference in the liquidity level of the two securities. Their model points out that liquidity supply is one of the main drivers of profitability within pairs trading. This is because pairs trading is a contrarian strategy: It tends to buy shares when most participants are selling, and sell shares when most are buying. Therefore, pairs traders are suppliers of liquidity to the market, and it is argued that this represents an important source of return to this strategy.

27.3.4 Sources of Risk and Return in Pairs Trading

Understanding the sources of risk and return is the most crucial part of analyzing any trading strategy. Even if a strategy has offered attractive returns in the past, managers and asset allocators must be aware of the sensitivities of the strategy to both systematic and idiosyncratic sources of uncertainty.

A pairs trading strategy is almost never a riskless strategy. As discussed in the following paragraphs, the strategy could experience losses. Available empirical evidence also supports this conclusion.[9] Therefore, it is also important to understand whether a typical pairs trading strategy has exposures to systematic and idiosyncratic sources of risk. If there are significant exposures to systematic sources of risk, then the strategy cannot be considered an arbitrage strategy and its return represents compensation for bearing those risks. More importantly, if exposures to systematic sources of risk are dominant, then the benefits of having a diversified portfolio of pairs trading strategies will be limited. On the other hand, if the idiosyncratic sources of risk are most important, then it will be essential for the manager to have a fully diversified portfolio of such strategies, as the rate of return from such a portfolio could come close to representing an arbitrage opportunity as the size of the portfolio increases.

Available empirical evidence indicates that there are both systematic and idiosyncratic risks present in a typical pairs trading strategy. The following sources of risk have been identified by previous studies:

- Noise traders risk: **Noise traders risk** is performance dispersion caused by idiosyncratic trading. In the same way that noise traders could create a spread between two co-integrated prices, they could cause the spread to widen even further. This would lead to losses, forcing the manager to close the positions before the two prices had even begun to converge. This risk is mostly idiosyncratic and thus could be diversified away.
- Fundamental risk: In this case, the spread between the two stocks could become permanent due to a fundamental change in the two firms. For example, one of the two firms could develop a new product that would change the fundamental relationship between the firms (e.g., Apple developing the iPhone and therefore becoming less connected to the computer industry). This risk is mostly idiosyncratic and thus could be diversified away.
- Corporate event risk: Corporate events such as mergers and spin-offs are most likely to change the relationship between the two stocks. This risk is a combination of idiosyncratic and systematic risks. Since industry mergers and acquisitions tend to be somewhat correlated, there is a common risk factor present in a pairs trading strategy that involves the same industry. To diversify this risk, it is important to use several pairs of stocks, with each pair coming from a different industry.
- Synchronization risk: **Synchronization risk** arises when market participants are slow to react to increased divergence between two stocks, leading the portfolio manager to consider closing the positions because the convergence has not taken place during a specific period. For example, empirical evidence has shown that speed of convergence for related stocks changes through time.[10] This indicates that this risk likely affects a large set of pairs of stocks and therefore may not be fully diversifiable. However, it is not clear if this represents a systematic source of risk in which the investor is compensated for bearing this risk.
- Liquidity risk: Stocks that are candidates for pairs trading must be scanned for sufficient liquidity. Since each strategy involves at least four transactions, bid-ask spreads and market impact could severely impact the performance of the strategy. Liquidity is not constant, and it tends to disappear during periods of financial stress. Therefore, this risk tends to be systematic and therefore is not

fully diversifiable. In fact, since pairs trading is a contrarian strategy, one could argue that the strategy compensates investors for providing liquidity to the rest of the market.

- SHORT-SALE RISK: **Short-sale risk** could arise if the investor is forced to cover the short positions because either the shares have been called in by the lenders or there is a short squeeze. In addition, the investor could face a short squeeze if a large number of investors are forced to cover their short positions. For example, in 2008, a short squeeze in shares of Volkswagen caused the stock price to increase by 500% in two trading days. A short squeeze tends to be firm specific and therefore can be diversified away.
- MODEL RISK: Since pairs trading is essentially a statistical arbitrage strategy, it relies on the assumption that the underlying models of stock returns and their estimated parameters are accurate and stable through time. To the degree that there are common flaws in the statistical models employed for different pairs, this risk could affect all the strategies being implemented by investors and therefore cannot be diversified away. However, this is not a systematic risk in the sense that the investor will not be compensated for bearing this risk.

Clearly, pairs trading strategies have exposures to some sources of systematic risk (e.g., liquidity), and therefore a portion of the return they generate can be considered compensation for bearing those risks. Empirical evidence cited previously argues that systematic risk can explain only part of the return. The unexplained part of the return is attributed to the presence of noise traders, complexity of the strategy, and leverage risk.

27.3.5 Pairs Trading Market Size and Historical Performance

Historically, equity market-neutral and statistical arbitrage strategies have occupied a relatively small portion of the hedge fund industry. HFR reports that as of the third quarter of 2015, dedicated equity market-neutral funds represented approximately \$52.6 billion, or 1.86% of total assets, up from \$40.8 billion in 2005. Though quite small in comparison to equity hedge funds—including long/short funds, which made up more than 28% of hedge fund assets—the equity market-neutral strategy serves as a diversifier and provides returns that are bondlike but with small exposures to equity and bond markets. Exhibit 27.12 displays the performance of the CISDM Equity Market Neutral Index as well as that of global equity and fixed-income indices for 2005–15.

27.4 CONCLUSION

This chapter has dealt with relative value arbitrage and equity market-neutral strategies, which entail constructing portfolios consisting of long and short positions in securities. It discussed how and why limits to arbitrage can create arbitrage opportunities that could be exploited by skilled managers. The chapter also described how the same forces that create arbitrage opportunities can be sources of risk in these strategies.

EXHIBIT 27.12 Risk-Return of Equity Market-Neutral Arbitrage Strategy

2005–15 (USD)	Annualized Mean Return	Annualized Std. Dev.	Maximum Drawdown	Information Ratio	Correlation with MSCI World	Correlation with Barclays U.S. Agg	Correlation with Barclays U.S. High Yield
CISDM Equity Market Neutral	5.54%	2%	−3%	2.59	0.48	−0.08	0.38
MSCI World	6.30%	16%	−56%	0.39	1.00	0.08	0.75
Barclays Global Aggregate	3.08%	6%	−10%	0.56	0.41	0.70	0.36
Barclays Global High Yield	7.34%	11%	−33%	0.68	0.80	0.27	0.98
Barclays U.S. Aggregate Bond	4.30%	3%	−4%	1.34	0.08	1.00	0.24
Barclays U.S. Corporate High Yield	6.91%	10%	−33%	0.68	0.75	0.24	1.00

Another important point highlighted by this chapter was that returns from these strategies may not be pure arbitrage profits. For example, convertible arbitrage strategies provide a service for issuing companies by increasing the supply of shares through shorting and eventually creating a permanent increase in the supply of those shares when the bonds are converted. Also, a portion of the return earned by pairs traders could be attributed to the service that they provide to the market. First, because it is a contrarian strategy, it tends to buy shares when most participants are selling, and sell when most are buying. Therefore, pairs traders are suppliers of liquidity to the market. Second, pairs traders benefit the whole market by forcing stocks to move toward their fundamental values.

NOTES

1. Shleifer and Vishny (1997) contains an early exposition of the idea of limits to arbitrage; for a comprehensive review of the academic literature in this area, please see Gromb and Vayanos (2010).
2. For further discussion, see L. H. Pedersen (2015).
3. Convertible arbitrage strategies were briefly discussed in Chapter 19 of the CAIA Level I book.
4. Ammanna, Kind, and Wilde (2003); Calamos (2003); Choi, Getmansky, and Tookes (2010).
5. For detailed discussion of convertible bond arbitrage and pricing of convertible bonds, see Calamos (2003).
6. This section is drawn from AIMA (2006).
7. For academic studies of pairs trading strategy, see Andrade, Di Pietro, and Seasholes (2005); Elliott, Van Der Hoek, and Malcolm (2005); Gatev, Goetzmann, and Rouwenhorst (2006).
8. For detailed discussion of pairs trading strategy, see Vidyamurthy (2004).
9. See Ehrman (2006); Gatev, Goetzmann, and Rouwenhorst (2006); Berg (2013); Engelberg, Gao, and Jagannathan (2009).
10. Do and Faff (2010).

REFERENCES

AIMA. 2006. “Convertible Bond Arbitrage.” AIMA Canada Strategy Paper Series. September 2006, Number 6.

Ammanna, M., A. Kind, and C. Wilde. 2003. “Are Convertible Bonds Underpriced? An Analysis of the French Market.” *Journal of Banking and Finance* 27 (4): 635–53.

Andrade, S., V. Di Pietro, and M. Seasholes. 2005. “Understanding the Profitability of Pairs Trading.” Working paper, Northwestern University.

Berg, F. 2013. “Return and Risk Exposure in Pairs Trading: Evidence from the German Equity Market.” Working paper, Copenhagen Business School.

Calamos, N. P. 2003. *Convertible Arbitrage: Insights and Techniques for Successful Hedging*. Hoboken, NJ: John Wiley & Sons.

Choi, D., M. Getmansky, and H. Tookes. 2010. “Convertible Bond Arbitrageurs as Suppliers of Capital.” *Review of Financial Studies* 23 (6): 2492–522.

Cox, J., S. Ross, and M. Rubinstein. 1979. “Option Pricing: A Simplified Approach.” *Journal of Financial Economics* 7: 229–263.

Do, B., and R. Faff. 2010. "Does Simple Pairs Trading Still Work?" *Financial Analysts Journal* 66 (4): 1–13.

Ehrman, D. S. 2006. *The Handbook of Pairs Trading: Strategies Using Equities, Options, and Futures*. Hoboken, NJ: John Wiley & Sons.

Elliott, R. J., J. Van Der Hoek, and W. P. Malcolm. 2005. "Pairs Trading." *Quantitative Finance* 5 (3): 271–76.

Engelberg, J., P. Gao, and R. Jagannathan. 2009. "An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity." Third Singapore International Conference on Finance.

Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst. 2006. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule." *Review of Financial Studies* 19 (3): 797–827.

Gromb, D., and D. Vayanos. 2010. "Limits of Arbitrage: The State of the Theory." NBER Working Paper 15821.

Pedersen, L. H. 2015. *Efficiently Inefficient*. Princeton, NJ: Princeton University Press.

Shleifer, A., and R. W. Vishny. 1997. "The Limits of Arbitrage." *Journal of Finance* 52 (1): 35–55.

Vidyamurthy, G. 2004. *Pairs Trading: Quantitative Methods and Analysis*. Wiley Finance. Hoboken, NJ: John Wiley & Sons.

CHAPTER 28

Hedge Funds: Directional Strategies

This chapter covers two discretionary directional hedge fund strategies: equity long/short and global macro. Given that the basics for each strategy have been covered in CAIA Level I, this chapter explores more advanced topics.

28.1 FINANCIAL ECONOMICS OF DIRECTIONAL STRATEGIES

The difficulty of predicting asset returns is well-known by both practitioners and academics. Attempting to predict future asset returns is arguably an impossible task in the extreme and a very difficult endeavor at minimum. The two discretionary directional strategies cited, equity long/short and global macro, are distinct from quantitative hedge fund strategies in that they attempt to take fewer bets with larger position sizes and much more concentrated positions, whereas the latter strategies tend to take more bets with smaller position sizes. In addition, quantitative strategies tend to be systematic and involve very little, if any, discretion.

Discretionary directional managers take fewer bets with larger position sizes because their approach is more labor intensive, and each investment opportunity needs to be analyzed carefully and separately. Therefore, information gathering and analysis aspects of the strategy make it difficult to build portfolios consisting of hundreds of positions. Quantitative approaches, in contrast, rely on computing power to apply their algorithms and models to large sets of investments rather quickly. However, since each individual investment has not been the subject of hours of analysis by the manager, it is prudent to diversify the risk and make smaller allocations. Discretionary managers rely heavily on information flow and their personal synthesis of such information incorporated into their own investment thesis. Many have argued that discretionary managers are thus much more skill-based than non-discretionary or systematic managers, as the alpha that discretionary managers attempt to extract is less obvious and not well publicized.

Equity long/short generally begins with a deep value-based and/or fundamental approach, similar to those pioneered by Graham and Dodd and Warren Buffett. Most managers have a very thoughtful understanding of the companies that they invest in. Additionally, global macro tends to be very "big picture" in nature. Having strong views on global interest rates, equity markets, and/or commodities is not uncommon for macro managers. Both managers would bet heavily on ideas based on strong convictions. Directional managers typically generate return streams with large volatilities, with the potential of large gains or losses each year.

28.1.1 Informational Market Efficiency

To better understand equity long/short and global macro strategies, consider these strategies from the context of the efficient market hypothesis, as described by Eugene Fama.[1] The **efficient market hypothesis (EMH)** states that securities markets incorporate all available information in the determination of the value of a security. The consequence of the EMH is that investors cannot achieve better risk-adjusted returns than could be achieved by holding the market portfolio. Here, the market portfolio is the aggregate supply of all assets, and many academics and practitioners proxy this "market" by the capitalization-weighted equity portfolio indices, such as the S&P 500 Index or the MSCI World Index. Others have argued that this market proxy has been misspecified, as it should include all possible investments, such as real estate and human capital.

In any case, Fama proposes three versions of the EMH: The weak form of the EMH states that asset prices fully reflect all past price information. Weak form efficiency would rule out the ability of technical analysis to give investors an edge over the market, as many technical analysts employ historical prices to generate technical trading signals, such as moving average crossovers, relative strength indicators, and Bollinger bands, that are used to trade the markets. Note that weak form efficiency does not include other publicly available information, such as balance sheet and income statements—just past prices. Next, the semistrong form of the EMH states that all publicly available information has already been incorporated into securities prices—past prices as well as any information in the public domain. Such information could be the fundamental information available on a regulatory agency's website (e.g., SEC.gov in the United States), as well as company press releases, earnings announcements, and information revealed on company earnings conference calls, to name but a few. If semistrong form efficiency were true, then the fundamental analysis whereby an analyst looks over publicly available financial information (e.g., financial statements, conference calls, press releases, and Twitter feeds) would not generate any advantage, since such publicly available information would be quickly incorporated into securities prices. Clearly, if markets were semistrong form efficient, we would not have a thriving equity long/short industry. Note that semistrong form efficiency leaves open the final kind of information: private information. To close this gap, Fama proposes the strong form of the EMH, which states that all information is reflected in securities prices. That is, both public and private information has already been correctly incorporated into securities prices. In practice, strong-form efficiency appears too extreme, as knowing private information on companies is deemed to be such an advantage that trading on such private information (and thereby causing prices to adjust) is outlawed by most regulatory agencies. Basically, trading on private information gives participants such an advantage that it is considered completely unfair. Even where insider trading on material nonpublic information is not considered to be illegal, CAIA candidates and members are required to avoid these activities according to the Code and Standards.

Every now and then we see news articles about greedy traders who are caught trading on private information. Additionally, we see the magnitude of such advantages in terms of the profits on those trades. Clearly, the markets are not strong form efficient, or we would not have such strict laws against trading on inside information, nor would we see such high returns from trading on inside information. Thus, we can rule out markets being strong form efficient.

Now let's consider semistrong form efficiency. Academics teach fundamental investing to most of the MBA students interested in learning finance. Students leave school armed with the ability to perform fundamental analysis using valuation models based on publicly available financials. They obtain jobs at hedge funds and mutual funds scouring the equity universe for undervalued stocks. Those who can identify valuable companies are rewarded with accolades and more assets to manage, thus generating more management fees for their business. So it may appear that there are advantages to finding and distilling such publicly available financial information, as businesses are willing to pay up for talented fundamental stock pickers. Although each year many try to gain an advantage by applying fundamental analysis on publicly available information in the quest to find good investments, such an endeavor is extremely difficult. Thus, the semistrong form of the EMH does not hold completely, probably less than the proponents of a semistrong form efficient market would like to concede. When someone does find some undervalued securities, it's clear that this advantage does not last for a long time. Undervalued securities are available in the markets, but once the community catches on that such a security is undervalued, other investors come in and bid up the security's prices. Markets rapidly adjust by incorporating publicly available information, causing such undervaluation to become properly priced.

Finally, are markets weak form efficient? Many technical analysts claim to have some advantages trading securities based purely on historical prices. While some may make a career out of such endeavors, it's clear that this is probably an even more difficult task to do vis-à-vis fundamental analysis. Clearly, many traders do know how to trade based on technical tools, but few will rely purely on their technical skills to sustain themselves. Thus, weak form efficiency appears to hold, on average, reasonably well within the markets. For a new approach to understanding market efficiency and active management see Pedersen (2015).

Equity long/short fundamental investing could be argued to be an affront on the semistrong form of the EMH. Equity long/short hedge fund managers are presumed to be better at finding and incorporating such publicly available information and identifying those securities that are undervalued. Additionally, since they can also short-sell securities, they can look for overvalued securities and take a short position accordingly. The hope would be that the market realizes this overvaluation, and prices of the security decline to reflect the poor fundamentals.

Terms such as "value trap" and "short squeeze" may hinder the ability to realize profits. A **value trap** occurs when a security is undervalued but may stay undervalued for a prolonged period. This long period of undervaluation may occur due to a variety of reasons, such as lack of new products, fierce potential competition, and ineffective management. Additionally, a short squeeze may occur when the fundamentals are extremely poor and many market participants agree that the stock price is overvalued and should go down, but in the near term, the stock price rises so quickly that many short sellers are forced to cover positions at a loss. Volkswagen was a good example, whereby the fundamentals in 2008 were horrible for the automobile industry, since demand was nonexistent for cars. However, Porsche successfully took revenge on the hedge funds that shorted Volkswagen. Porsche disclosed that it owned 42.6% of the stock and had options to acquire another 31.5%. As it revealed this information, market participants who had shorted Volkswagen realized that they had to cover their shorts. In one day, Volkswagen's price made it the largest stock in the world by market capitalization.[2] Although the fundamentals to Volkswagen's business were deteriorating during that time, the short squeeze by Porsche significantly impacted those fundamental equity long/short managers.

Some of the most influential attacks on the EMH come from the many empirical anomalies that have been discovered and published over the years. Harvey, Liu, and Zhu (2016) propose that academics should employ a higher standard given the extensive data mining that has occurred over time. Instead of the traditional *t*-statistics greater than 2.0 test, they state that new discoveries should overcome a much higher hurdle rate of 3.0. Moreover, they make a very controversial statement: "We argue that most claimed research findings in financial economics are likely false."

28.1.2 Behavioral Finance

Many critics challenge the basic assumptions of the EMH. The premise that investors are rational and value securities without any biases has come under attack by many behavioral financial economists. Behavioral finance is built on two main building blocks: (1) limits to arbitrage, which states that rationally behaving traders may have difficulty undoing dislocations that are caused by less rational traders, and (2) **cognitive psychology**, which is a broad category that attempts to capture the many types of deviations from full rationality. Basically, it attempts to understand how people think. Emotions have been shown by many to affect investors' decision-making abilities. For example, market panics are events in which fear selling and herding behaviors often occur. Some other examples include leverage aversion, anchoring, and dispositions. The following pages present a few of the more popular behavioral biases as well as recent research that attempts to better understand such biases.

It should be noted that hedge funds see such periods of irrationally behaving markets and less rational market participants as ideal periods or situations in which to capitalize on such behavior. The birth of the behavioral finance industry has significantly contributed to our understanding of security price behavior as well as the reasons that prices may deviate from fundamental value. Barberis and Thaler (2003) provide an excellent literature review on this topic.

Leverage aversion/restriction. Frazzini and Pedersen (2013) formulate the theory of leverage aversion and provide strong empirical support for it. Their leverage aversion theory predicts that leverage-constrained investors bid up high-beta assets. A portfolio containing many of these high-beta assets would be a substitute for a portfolio containing average-beta assets but levered up to create a high-beta portfolio. Such high-beta portfolios are attractive to fund managers who have strong beliefs that market prices will be rising, and to fund managers whose performance lags the overall market and who see investing in a high-beta portfolio as the only means of catching up with their peers and the overall market. As a result, prices of high-beta stocks tend to be bid up and, therefore, are associated with negative alphas. A strategy of **betting against beta** (BAB) occurs when one takes long positions in low-beta assets and short positions in high-beta assets in anticipation of significant positive risk-adjusted performance. By examining historical returns generated from BAB across the U.S. equity market, Frazzini and Pedersen show compelling evidence of statistically significant outperformance: BAB generated 70 basis points per month over the period from January 1926 to March 2012. Additionally, they show the robustness of their results by examining international equity markets, Treasury bonds, corporate bonds, and futures. Further studies have expanded the concept of leverage aversion and conclude that the predictions of modern portfolio theory should be altered. They show that safer assets should generate higher risk-adjusted returns than riskier assets. Thus, leveraging safer assets allows investors to be on the right-hand side of the tangency portfolio—that is, to hold a portfolio with the same level of risk and higher average returns.

More recent research examines Warren Buffett's Berkshire Hathaway and concludes that Berkshire has a significant alpha to traditional risk factors. However, once the researchers control for exposure to BAB and quality minus junk (QMJ), the alpha attributable to the Oracle of Omaha becomes insignificant. Asness, Frazzini, and Pedersen (2014) define **high-quality assets** as securities that are safe, profitable, growing, and well managed. They find that a portfolio constructed of long high-quality stocks and short low-quality stocks generates significant risk-adjusted returns in the United States as well as in 23 other countries. They estimate that the cheap, safe, and low-beta stocks in the portfolios can be enhanced by leverage.

Sentiment sensitivity. Baker and Wurgler (2006, 2007) review the literature on sentiment and stock markets. **Sentiment** is broadly defined as beliefs about future cash flows and risks that are not justified by an objective analysis of the facts. The authors explain that the level of stock prices in the aggregate depends on sentiment. In particular, stocks of small-capitalization, younger, unprofitable, high-volatility, and non-dividend-paying growth companies, or stocks of firms in financial distress, are likely to be disproportionately sensitive to broad waves of investor sentiment. They argue that betting against investor sentiment may be a very costly and risky proposition, as rational investors may not be aggressive enough to force prices back to their fundamental values. Thus, there are limits to arbitrage in the presence of sentimental investors (noise traders). For a variety of reasons, these investors may become too optimistic or pessimistic about certain sectors of the market or the overall market and cause prices to deviate from their fundamentals. Some examples include the Internet bubble and burst of 1999–2000 and the real estate bubble and burst of 2006–9. The authors construct a sentiment index and show that the behavior of the index has a significant relationship with the behavior of certain stocks, such as those that have small capitalizations and are younger, unprofitable, highly volatile, or non-dividend-paying, or stocks in financial distress. Given the higher transaction costs and trading risks, they argue that these stocks are harder to arbitrage, and valuation mistakes are more prevalent. They incorporate six underlying proxies for their aggregate sentiment index: (1) the discounts on closed-end funds, (2) the turnover of New York Stock Exchange (NYSE) shares, (3) the number of initial public offerings (IPOs), (4) the average first-day returns on IPOs, (5) the equity share in new issues, and (6) the **dividend premium**, which is defined as the difference between the average market-to-book-value ratios of dividend payers and nonpayers.

There have been other approaches to the construction of sentiment indices, and studies have shown that these indices may explain the behavior of certain stocks through business cycles. For example, let's consider the behavior of high-beta versus low-beta stocks and the behavior of high-accrual versus low-accrual stocks. As discussed in the CAIA Level I book, the CAPM predicts that high-beta stocks should earn higher rates of return than low-beta stocks. Also, the accrual anomaly predicts that stocks of companies with high accruals (i.e., companies in which reported earnings are higher than reported cash flows) tend to underperform stocks of companies with low accruals (i.e., companies in which reported earnings are less than reported cash flows). Using indices of investor sentiment, studies have shown that predictions of the CAPM and accrual anomaly in poor-sentiment periods hold up well. That is, high-beta stocks outperform low-beta stocks, and low-accrual stocks outperform high-accrual stocks. However, in positive-sentiment periods, these predictions are not supported by data. Basically, it appears that in low-sentiment periods, investors are behaving rather rationally with regard to the capital asset pricing model (CAPM) and the accrual anomaly. In strong-sentiment environments, we see possible euphoria taking over and participants overpaying for high-beta and high-accrual companies. Furthermore, in high-sentiment periods, it seems as though market participants are relying more and more on the analysts' earnings revisions, since CAPM and fundamentals, like accruals, are less reliable.

Digital and social media have been used to create measures of investor sentiment. For example, Liew and Wang (2015) created a sentiment index based on the number of tweets that were made about impending IPOs, showing that Twitter sentiment predicts the cross section of IPO returns. By breaking down the first-day performance of IPOs using the opening price and open-to-close return of first-day trading, they identify both a positive predictive relationship and a negative predictive relationship, respectively. They conclude by stating, "We document that Tweet sentiment matters for IPO first-day performances but the nature of this relationship appears very complex."

Overconfidence. "How good are you at investing?" When asked such a question about investment skill and knowledge, many investors are inclined to attribute to themselves superior skills. Unfortunately, it has been shown that people are poor at calibrating probabilities and confidence intervals. Alpert and Raiffa (1982) show that 98% confidence intervals include the true quantities only about 60% of the time. Fischhoff, Slovic, and Lichtenstein (1977) document miscalibration of probabilities. They find that events that people think are certain to occur only occur only 80% of the time. Finally, many argue that overconfidence leads investors to overtrade their portfolios, incurring heavy transaction costs and drags to performance. Barber and Odean (2001) examine the trading activities in discount brokerage accounts and find that the more someone trades, the worse the person performs on average. Interestingly enough, they find that men trade more and thus do worse than women.

Anchoring effects. Anchoring occurs when a person is biased due to prior views and cannot properly integrate new information. Studies have shown that the cognitive bias of anchoring is present in sell-side analysts' forecasts of earnings. They document that analysts make optimistic forecasts when the forecasted earnings per share (EPS) are lower than the industry median. In other words, industry-median EPS estimates serve as an anchor, and therefore earnings estimates of poorly performing companies tend to be too high and close to the industry's median. On the other hand, earnings estimates of firms that are performing well tend to be underestimated. As a result, we observe that the future stock returns are significantly higher than forecasted for firms with above-median EPS forecasts. Additional studies have found that expert consensus forecasts of monthly economic releases from the Money Market Services surveys have a systematic bias toward the previous month's data. Given that forecasts of any given release are anchored to prior recent months' values, the studies argue that this bias gives rise to predictable surprises.

CONFIRMATION BIAS. Confirmation bias occurs when one selectively employs evidence that supports a given claim or belief and minimizes contradictory evidence (aka "selective hearing"). Duong, Pescetto, and Santamaria (2010) directly test for confirmation bias in the context of value versus glamour investing and fundamental analysis in the United Kingdom. Value stocks are defined as those with a high fundamentals-to-price ratio (high book-to-market ratio, earnings-to-price ratio, and cash-flow-to-price ratios); alternatively, glamour stocks are those that have a low fundamentals-to-price ratio. Using sample data consisting of UK stocks from 1991 to 2007, the researchers document investors' asymmetric reactions to both good and bad news. Value investors underreact to good news and process bad news overconfidently, whereas glamour investors underreact to bad news and process good news overconfidently. In other words, each type of investor reacts more strongly to the news that confirms the group's prior beliefs.

LOSS AVERSION/DISPOSITION EFFECT. **Loss aversion/disposition effect** captures the notion that investors typically prefer to avoid losses rather than acquire gains. Kahneman and Tversky (1979) critiqued the expected utility theory and developed their alternative prospect theory. In prospect theory, agents underweight those outcomes that are probable vis-à-vis those outcomes that are certain. Such tendencies contribute to the risk aversion in choices involving sure gains and the risk-seeking preference in choices involving sure losses. One example is that many investors lock in small profits and fail to cut off losses quickly. The adage "Let your winners run and cut your losses" is a constant reminder to investors of their potential miscalculations due to loss aversion.

While it might be difficult to create specific trading strategies that can exploit deviations from fundamentals caused by these behavioral biases, at least they demonstrate that investors may not act as the strict rational agents required by the EMH. Further, it might be possible to devise trading strategies that would benefit from the collective impact of these behavioral biases.

28.2 EQUITY LONG/SHORT

Fundamental equity long/short, typically referred to as long/short, is an investment strategy associated with hedge funds whose managers buy equities that are expected to rise in value and sell equities that are expected to fall in value. This strategy is typically implemented through fundamental stock selection and, to a lesser extent, by varying total net exposure, also known as market timing. Note that, in practice, the classification of a long/short hedge fund is somewhat subjective, and therefore caution should be used when a hedge fund is categorized as such.

Equity hedge fund strategies may be differentiated by the extent to which positions are determined using discretionary techniques rather than systematic techniques. Discretionary techniques usually rely on fundamental analysis, whereas systematic techniques rely more on technical analysis. Furthermore, discretionary techniques often depend on qualitative analysis, whereas systematic techniques depend more on quantitative analysis. This chapter provides an in-depth review of fundamental and qualitative equity hedge fund strategies.

Common equity hedge fund strategies are short bias, equity market-neutral, long/short, and leveraged strategies, such as 130/30 funds. This chapter focuses on fundamental equity hedge fund strategies in the context of equity long/short strategies, although fundamental strategies can be deployed throughout the directional and leverage spectrums of equity hedge funds.

28.2.1 Overview of Investment Opportunities

Since equity long/short managers typically invest in stocks, the investment opportunity set is all possible combinations of stocks. For simplicity, this discussion excludes derivatives and any private investments, though some managers may employ derivatives. Given that long/short managers can buy and sell short, all combinations of long and short stock positions need to be included. Additionally, long/short managers can be underinvested and have a positive allocation to cash, or be overinvested by borrowing cash to achieve leverage. The possible number of combinations grows very large when we consider the continuum of allocation choices across long stocks, short stocks, and cash/leverage. Moreover, long/short managers can vary their exposures across time by increasing and decreasing total leverage while holding the portfolio's composition constant, by changing the composition of the portfolio while holding the leverage constant, or by engaging in a combination of both.

Equity long/short portfolios are typically highly concentrated in a relatively small number of stocks, with core positions ranging from three to 10 stocks, and noncore positions ranging from 20 to 40 stocks. This is much more concentrated than equity market-neutral or statistical arbitrage portfolios, which may have hundreds or even thousands of positions. Long/short managers typically have much lower turnover, with longer holding periods than those of equity market-neutral or statistical arbitrage managers. Long/short managers may hold positions for many years. It is not uncommon to hear that the investment horizon for these managers is on the order of three to five years.

28.2.2 Value, Growth, and Blend Approaches

Long/short managers vary concerning investment style classifications analogously to how mutual funds differ in terms of styles, such as value, growth, and blend. **Value long/short managers** employ traditional valuation metrics, such as the book-to-market ratio, earnings-to-price (E/P) ratio, dividend yield, and the ratio of P/E to earnings growth rate (i.e., the PEG ratio), to look for undervalued companies. Managers following this approach tend to be contrarians and to invest in companies that are temporarily out of favor. Empirical evidence shows that value stocks tend to outperform growth stocks through many business cycles.[3]

The **growth approach** to fundamental long/short equity investing is to overweight companies perceived as having higher potential to deliver large increases in revenues, earnings, and/or cash flows. Long/short growth managers are attracted by top-line growth numbers and are willing to look past weak current earnings in the presence of aggressive sales growth. Often they invest in small high-tech companies because large companies in mature industries generally lack the same growth opportunities. Paying a reasonable price for growth companies is known as the growth at a reasonable price (GARP) approach.

The blended approach to fundamental long/short equity investing employs both approaches. Managers vary their investment process depending on the macro environment. Value stocks tend to have low betas relative to growth stocks. When anticipating down markets, managers may take a value investment approach, but in anticipation of rising markets, they increase their allocations to growth companies, thereby blending the two approaches.

28.2.3 Bottom-Up Approach versus Top-Down Approach

Fundamental equity long/short managers can also be classified as being either bottom-up or top-down investors. Most fundamental equity long/short managers are bottom-up investors—essentially stock pickers who tend to have concentrated portfolios. These investors are less concerned with market timing and forecasting macroeconomic trends or relative industry performance. Rather, they are concerned with opportunities and threats faced by individual companies and subsequently focus on their strengths and weaknesses. This framework is referred to as a **SWOT analysis**, in which an investment analysis is driven by four categories: strengths, weaknesses, opportunities, and threats.

The goal of a **bottom-up fundamental analysis** is to estimate the value of a company's stock based on firm-level forecasted sales, expenses, and earnings. These forecasts provide an estimated cash flow stream, which is discounted to arrive at the value of the company. The equity value is found after subtracting the value of the company's debt. Given that forecasting the future is challenging at best, analysts may generate a set of valuations based on good, medium, and bad scenarios, and then use the weighted average of these valuations where they correspond to the probability of the scenarios occurring. An example of a good scenario could be assuming high growth in company sales; a bad scenario could be assuming low growth.[4]

It is common for bottom-up long/short managers to spend substantial time away from their offices checking distribution channels, evaluating production lines, examining traffic at retail stores, and so on. Bottom-up managers often have researchers on the ground ferreting out any useful information that is not included in public documents. They perform detailed due diligence on the companies they hold and in which they intend to invest. For example, managers may ask industry experts about their views on a company's products and processes, competitive advantages, patents, and political and regulatory risks. Typically, this strategy focuses on companies that have limited analyst coverage, based on the idea that limited competition for information may indicate investment opportunities in which public information is not fully reflected in the market price of the equity. Bottom-up managers build their portfolios one stock at a time, focusing on characteristics of single stocks rather than a broad economic theme or portfolio-level targets.

Some long/short managers apply a **top-down fundamental analysis**, in which a company's perceived value is driven by a few broad investment and macroeconomic themes. Managers seek to forecast macroeconomic forces that would drive a sector's return, and implement their views through diversified portfolios and exchange-traded funds (ETFs). They tend to have strong views on the current stage of the business cycle, inflationary expectations, and monetary and fiscal policies.

Influential economy drivers are important to top-down managers, who spend less time on company-specific analysis. Following are some examples of the types of inquiries these managers may make:

- If the euro disintegrates, how will equity markets react?
- If gold is trading at an all-time high, will increased supply or reduced demand cause a reversal?
- What impact might Chinese government actions have on the economies of other nations?
- Are nations such as Brazil, Russia, India, and China going to fuel global economic growth?
- If oil prices rise or fall substantially, what are the implications for equity markets and various industries?
- What is the impact of expansionary monetary policy on global financial institutions?

Top-down long/short managers have a strong understanding of macroeconomic forces and their impact on financial markets, and they understand various interactions among different segments of capital markets: lead-lag relationships between the fixed-income, currency, commodity, real estate, and stock market sectors.

28.2.4 Fundamental Equity Valuation Models

Two very popular approaches that are well-known among bottom-up investors are Gordon's growth model and the enterprise valuation model.

Let us consider **Gordon's growth model (GGM)**, or the **perpetual growth model**, in which a stock's value is estimated as the present value of a perpetual stream of dividends assumed to grow at a constant rate. As an investor in a given stock, the stream of income that you would receive if you held the stock in perpetuity would be all of its future dividend payments. Thus, it would be reasonable to argue that the value of a stock should be the present value of the expected stream of future dividend payments discounted appropriately. For example, if V_0 is the value of the stock today and div_t is the dividend payment at the end of time period t, and if we assume that the appropriate discount rate is k_j for the dividend paid at the end each of period, then the value of a stock is as follows:

$$V_0 = \sum_{t=1}^{\infty} \frac{E[div_t]}{(1+k)^t} \tag{28.1}$$

Note the $E[\ ]$ is the expectation operator, since dividends could be unknown today. To make this equation easier to solve, assumptions are often made about uncertain future dividend payments and discount rates. For example, if we assume that the first dividend payment of div is known and that once it is paid, the dividends grow at the rate of g in each consecutive period, then the perpetuity with growth formula can be employed as follows:

$$V_0 = \frac{div_1}{k-g}. \tag{28.2}$$

Suppose a stock pays \$1 million in total dividends each year in perpetuity, and the appropriate discount rate is 10% per year. The present value of this stock would be \$1 million/0.10 = \$10 million. Now if we assumed that after the first dividend of \$1 million the company would continually grow its dividend at 5% per year, then the present value of the company would be \$1 million/(0.10 – 0.05) = \$20 million. If we further assume that this stock is trading at a market capitalization of \$3 million, then according to Gordon's growth model, the price (\$3 million) compared to value (\$10 million/\$20 million) would be a bargain in either scenario.

It is important to use caution when applying Gordon's growth model, since the valuation hinges on the assumptions of the future stream of dividends and the discount rates. Consider how the valuation of such stock would change if market conditions changed drastically. For example, if in the third year the company could no longer pay out any dividends, or if the fundamental risk of the industry and the company were increased substantially, what would happen? If dividends decrease to zero, then the value of the company would consequently decrease. If the discount rate increases, the value of the company would consequently decrease. Clearly, the company's valuation would have to be recomputed given the new conditions. It is also difficult to apply the GGM to stocks that do not currently pay dividends, as we would need to speculate on the size and timing of the first dividend payment. The GGM is known to conservatively value companies, as the model gives a firm a zero value if it is assumed that dividends will not be paid.

The enterprise valuation model attempts to calculate the total value of a firm and then extracts the value of the firm's equity by noticing that the value of the firm should equal the sum of equity and debt values. In particular, the enterprise value of a firm is defined as the market value of its operating assets:

$$\text{Total Value of Assets} = \text{Enterprise Value} + \text{Cash} \tag{28.3}$$

$$\text{Total Value of Assets} = \text{Debt} + \text{Equity} \tag{28.4}$$

$$\text{Equity} = \text{Enterprise Value} + \text{Cash} - \text{Debt} \tag{28.5}$$

The market value of a firm's operating assets is calculated using the free cash flow generated by the firm. **Free cash flow to the firm (FCFF)** is the total cash flow that is available for distribution to shareholders and bondholders of the firm. The FCFF is considered to be available for distribution to all suppliers of capital, because the cash needed to support the firm's operations as well as investments is already accounted for. In particular, FCFF is defined as follows:

$$\begin{aligned}\text{FCFF} = {} & \text{Net Income} + \text{Noncash Charges} + \text{Interest Expense} \times (1 - \text{Tax Rate}) \\ & - \text{Investments in Fixed Assets and Working Capital}\end{aligned} \tag{28.6}$$

Noncash charges refer to expenses that do not represent any cash payment. Depreciation charge is an important example of a noncash charge. The after-tax cost of interest is added back to net income because we want to calculate the cash that is available for distribution to both shareholders and bondholders. However, because we need to account for the tax benefits that interest expense provides, the after-tax value of interest expense is added back. Finally, we need to account for the cash that is used to make fixed investments or working capital investments. For example, cash may be used to acquire equipment and machinery (fixed-asset investment), to purchase raw material (working capital), or to provide credit to customers (working capital).

EXAMPLE: The 2014 net income available to shareholders of Digital Corporation is \$400 million. Annual depreciation and interest expense are \$40 million and \$30 million, respectively. The firm invests \$10 million in new equipment during 2014 and increases its inventory by \$5 million. The marginal tax rate for Digital Corporation is 30%. What is the FCFF of Digital Corporation?

$$\text{FCFF} = \$400 + \$40 + \$30 \times (1 - 0.3) - \$10 - \$5 = \$446 \text{ million}$$

To obtain the enterprise value (EV) of a firm using FCFF, we need to discount the estimated future values of FCFF using the firm's weighted average cost of capital (WACC). That is,

$$EV_0 = \sum_{t=1}^{\infty} \frac{FCFF_t}{(1 + WACC)^t} \tag{28.7}$$

The weighted average cost of capital, as the name implies, is the weighted average of the equity cost of capital, typically determined by the CAPM and the after-tax rate of interest the firm pays on its debt.

EXAMPLE: Suppose the FCFF of Digital Corporation is expected to grow by 7% per year going forward. What should be the total value of Digital Corporation if the WACC is 12.84%?

$$EV_0 = \frac{\$446 \text{ million}}{0.1284 - 0.07} = \$7{,}637 \text{ million}$$

Once the EV of the firm is estimated, the value of equity can be obtained by adding the value of its cash and subtracting the market value of its debt.

EXAMPLE: Digital Corporation holds \$50 million in cash, and the total value of the company's debt is \$500 million. What should be the value of its equity?

$$\text{Value of Firm} = \text{EV} + \text{Cash} = \text{Value of Equity} + \text{Value of Debt}$$
$$\$7{,}637 + \$50 = \text{Value of Equity} + \$500$$
$$\text{Value of Equity} = \$7{,}187 \text{ million}$$

28.2.5 Sector-Specific and Activist Investment Approaches

Additionally, long/short managers may be broken down into two groups: generalists and sector specialists. **Generalists** invest across a wide universe of stocks, whereas **sector specialists** tend to invest within a specific equity sector, such as financials, health care, or technology. Additionally, generalists may focus on a particular country or region. Some emerging market long/short hedge fund managers may market themselves as emerging country specialists with a local market presence. This would mean that they have analysts within a given emerging market country who know the local dialect, have an extensive network into companies, and may even have a deep understanding of the market's capital structure and dynamics. Sector funds are typically run by sector specialists who at one time were sector analysts for investment banks and brokerage firms. Most of these managers have been following a group of stocks within their industry specialization for many years. Portfolios are constructed by buying the stronger players within a given sector and selling the weaker ones. Popular sector funds focus on areas in which highly specialized skills are necessary, such as finance, health care, biotech, technology, real estate, and energy. In addition to the generalist/sector specialist distinction, long/short managers employ various strategies, such as sector momentum, day trading, and market timing.

Activists are long/short managers who take a very public stance on their investments. Activist strategies rely on corporate governance changes to unlock value. Activists are open about confronting and criticizing senior management and boards of directors of public companies. Research has shown that they are quite effective at changing the composition of a company's board of directors, and they generally get management to adopt recommended changes to the corporate structure. Corporate governance changes can favor one group of stakeholders over another. For example, an activist manager may recommend increasing dividends, benefiting shareholders to the detriment of bondholders.

A study by Brav et al. (2008) reports that between 2001 and 2006, activist hedge funds in the United States proposed strategic, operational, and financial remedies to the firms in which they had taken significant positions, and attained success or partial success in two-thirds of the cases studied. The excess (abnormal) stock returns experienced by the target firms upon the announcement of a proposal were approximately 7%, with no reversal during the subsequent year. Furthermore, it was reported that the target firms experienced increases in payout, operating performance, and higher CEO turnover after activism.

28.2.6 Mechanics of the Long/Short Strategy

This section discusses the mechanics of the traditional long/short investment process. Though the steps taken by any one manager will vary, essentially all long/short managers begin with an investment idea and end with portfolio risk management.

28.2.6.1 Idea Generation The first and by far the most critical step is to generate good investment ideas. Some managers screen the universe of stocks based on fundamental ratios or technical indicators so as to reduce the total number of stocks to a manageable size. Others read industry newsletters, research reports, market commentary, academic research, or other written sources of information to gain investment insights. Additionally, some managers attend investment conferences, trade conferences, and idea luncheons or dinners to develop new ideas. The value of a solid network of colleagues cannot be overestimated in uncovering and refining new ideas. Some may talk to friendly CEOs or CFOs, while others may scrutinize forms required by regulators, such as Form 13F reports, for ideas. Any source that is publicly available can be employed for idea generation.

28.2.6.2 Optimal Idea Expression The next step, instinctive to seasoned long/short managers, involves deciding how to best express an idea—in other words, determining the best investment decisions that can be made based on the investment idea. During this process, the manager may make the following inquiries:

- What trade should be executed to extract the highest return from the idea?
- What price level will either confirm or negate the idea's validity?
- Will the stock price move straight up and become range-bound, or will it move slowly upward over many months?

For example, suppose a manager believes that the soon-to-be-released earnings of Company XYZ will exceed the consensus forecast. The manager must decide whether to buy the stock, buy call options, sell put options, or express the trade through sector ETFs. In the interest of executing the optimal trade, the manager must further consider the downside risk and the time frame.

28.2.6.3 Sizing the Position The next step works in conjunction with the previous step. Typically, a long/short manager has existing positions, so she needs to understand how the new position will fit within the context of the current portfolio. Sizing the position may require resizing the other opportunities within the portfolio. Positions are generally sized according to the level of the manager's conviction regarding the idea. If the long/short manager's conviction is high, she is likely to take a large position. If her conviction is low, she may build a toehold position and wait to see whether the idea is a good one. She may then increase the size of the position if her conviction increases.

28.2.6.4 Executing the Trade Many long/short managers will execute their own trades. That is, they will physically place the trade into the market via an electronic trading system or through their broker. The majority of long/short managers have extensive trading experience. They also tend to have a strong understanding of intraday technical indicators and have experience in determining when the technical environment is favorable for entering or exiting positions. When executing their trades, they need to consider whether they should buy aggressively (lift offers), sell aggressively (hit bids), or trade passively (join or improve bids or offers). Other important considerations when executing trades include the following:

- The liquidity of the underlying security at the time of execution
- Whether there is a major announcement due out that may move the market
- The availability of the stock to be borrowed in the case of a short sale

Successful managers should possess a keen sense of the trade's potential market impact, which may increase trading costs. If a manager has a large order and tries to execute it all at once, he may move the market so violently that the ultimate average price paid would be very expensive. However, if he splits the order into smaller pieces or executes at either a time-weighted average price (TWAP) or a volume-weighted average price (VWAP) throughout the day, he may achieve a better average price. The trade-off between alpha decay and order completion should be evaluated by determining what opportunity costs arise by waiting to execute. If the manager executes quickly, how much does the trade influence the market? For instance, the impact would be much greater for thinly traded stocks than for liquid stocks. Additionally, long/short managers should have a firm understanding of the costs of commissions and exchange fees. Upon execution of the trade, long/short managers will usually examine their slippage reports (either generated internally or provided by the prime broker), which show the difference between the price traded and the price anticipated before the trade. Some trades may not be fully executed, requiring long/short managers to deal with the consequences. The urgency of trading informs the trader of the execution strategy. Higher turnover strategies with shorter holding periods, such as momentum trades or investing for an upcoming earnings report, require faster execution. Managers of strategies with longer holding periods, such as those used by value investors, may seek to trade more patiently and wait for lower prices.

Fund managers—or, more particularly, head traders—will often examine the average daily volume of each stock in an attempt to assess the underlying liquidity of each position. Some constraints may be enforced (such as never having positions larger than X% of the average daily trading volume) to ensure the ability to exit the position quickly when necessary. This average may be computed over periods of various lengths, known as rolling window analysis. However, such attempts at managing portfolio liquidity become increasingly challenging as the manager's assets under management (AUM) increase, typically due to successful performance and fundraising. As the AUM increases within a given long/short fund, the slow creep into larger-capitalization stocks in search of higher liquidity may become more and more difficult to avoid, which may ultimately reduce the trading opportunities and alpha of the fund.

28.2.7 Understanding and Managing the Risks of Long/Short Investing

Once the order has been executed and the confirmations are checked, portfolio positions are examined closely to see whether the investment idea is working. Corporate actions—such as merger announcements, spin-offs, and divestitures—require monitoring because they may change the risk profile of the company. Many long/short managers work with risk managers who go well beyond simply reporting net and gross exposure of the fund. Many employ sophisticated risk models from various risk vendors, such as Barra, Northfield, and Advanced Portfolio Technologies (APT), which assist fund managers with the estimation and daily reporting of factor exposures. Additionally, standard risk reports are typically generated using value at risk (VaR) analysis, in which the portfolio and all of the various ways it can be sliced—longs, shorts, sectors, regions, liquidity buckets—are examined. Furthermore, scenario analysis can be helpful in understanding the behavior of the portfolio in stressed markets (e.g., what if this portfolio had been held during the tech crash of 2000?). Some of the smaller long/short managers may default to using their prime broker's risk reports, since hiring a seasoned risk manager is expensive.

A brief commentary on the short position in a long/short portfolio is in order. Though short selling is a defining characteristic of a successful long/short manager, long/short managers typically run net long. Empirical research has shown that, on average, long/short managers have positive net exposures to the overall market, with a market beta around 0.5.

Theoretically, a short position may lead to unlimited losses. Additionally, when a short position goes against the manager, it is far more painful than when a long position goes against the manager. This is best illustrated in the following example.

Suppose a long/short manager has two investments, A and B, and she holds a long position of \$100 in A and a short position of \$100 in B. If both positions go against her by 10% in one day (A goes down from \$100 to \$90, and B goes up from \$100 to \$110), the manager loses \$10 on the long A position and \$10 on the short B position. Now, if both positions go against her the next day by another 10%, she will lose only \$9 on A, as A goes down from \$90 to \$81, but she will lose \$11 on B, as B goes up from \$110 to \$121, even though the percentage changes are both 10%.

Notice that the dollar losses in a long position become mitigated as the price decreases, whereas the dollar losses in a short position become magnified as the price increases. This subtle difference makes short selling stocks riskier than buying stocks.

Long/short managers pay much more attention to their short positions than they do to their long positions. The process of short selling is very different from purchasing stock. In many countries, short sales can be executed only on an uptick. This makes execution much more difficult and market impact much larger when stock prices are falling, as the stock may tick lower for a number of trades before the first small increase in price occurs, at which time the stock can be sold short. Stock must also be borrowed from an owner of the stock before it can be sold short. The short seller may be forced to cover the stock if the owner (lender) of the borrowed stock sells the stock, which must be delivered to the new owner. Short squeezes are another unique risk of short selling, in which the stock price rises rapidly and short sellers are forced to cover their positions for risk-management purposes. In heavily borrowed stocks, short squeezes are rumored to result from a concerted effort by some stockholders who would like to see the stock move sharply higher. Those stockholders buy the stock rapidly, hoping to maximize the market impact of their trades and to force the short sellers to cover their positions and add to the buying pressure, driving the stock significantly higher.

Other risks of short selling include regulatory measures aimed at limiting or preventing short sales.

28.2.8 Managerial Expertise and Sources of Returns

Two closely related issues in analyzing fundamental equity hedge fund strategies such as long/short are managerial expertise and sources of returns.

28.2.8.1 Expertise of Equity Long/Short Fund Managers Fundamental equity hedge fund managers should have an intimate knowledge of their stocks. For each stock within their portfolios, they are expected to know the company's business model, revenue generators, costs and expenses, product lines, comparative advantages, market opportunities, competition, and recent changes to the corporate structure. In addition, they should be aware of the company's weaknesses, threats to its business model, and new threats in technology and possible innovations. In establishing long positions, equity long/short managers look for solid companies with a defendable competitive advantage that are trading at a discount to estimated fair value. It is common to hear long/short managers describe their strategy as "value with a catalyst" or "growth at a reasonable price."

Fundamental equity long/short managers are well trained in fundamental analysis, often with previous experience at large hedge funds or as research analysts at top investment banks. Some long/short managers may come from long-only mutual fund complexes or successful private equity shops. Regardless, these managers must be well equipped to analyze balance sheets and income statements and to make projections about a company's future earnings prospects to build discounted free cash flow models. These models are often updated daily or at least weekly.

Many fundamental equity hedge fund managers are strong value investors, who, like Warren Buffett, look to invest in real businesses. These managers are often well versed in analyzing income statements and balance sheets, as well as understanding their associated footnotes. They listen to company conference calls, perform independent research, and have a strong grasp of what it takes to run a successful company. Some have strong accounting backgrounds, and others have law degrees, but the common skill among these long/short managers is that they can analyze a business in depth. Once they understand a business, they put together a forecasting model that predicts future earnings and cash flows. They value the company by determining an appropriate rate (commonly using the weighted average cost of capital) to discount projected future cash flows. They then compare this valuation with the market's assessment of the company. If the market's assessment is lower, the manager will buy the stock; if the assessment is higher, the manager will short it. A very important advantage for long/short managers is to have access to timely information through a strong network of contacts.

Large-capitalization stocks are well covered by equity analysts, making mispricings less likely. As a result, value-oriented equity long/short managers tend to focus their research on small-cap and mid-cap companies. This is consistent with the empirical evidence indicating that these managers have a positive bias toward small-cap stocks. Additionally, lesser-known names may have more inefficiencies when it comes to their potential misvaluations, which is consistent with the informational efficiency theories of equilibrium-based models conditional on the participation of agents.

28.2.8.2 Sources of Returns from a Fundamental Long/Short Strategy One potential source of returns is firm-specific informational inefficiency. If capital markets are informationally inefficient at the semistrong level, then the fundamental analysis may generate superior risk-adjusted returns by establishing long positions in underpriced equities and short positions in overpriced equities. Equity long/short hedge fund managers may attempt to exploit informational inefficiencies either by locating information not generally used by other investors or by being better at evaluating information that is commonly considered.

Fundamental equity hedge fund managers tend to be value-focused and, as discussed previously, are drawn to smaller stocks and equities that attract less attention from large institutional investors. This empirical observation is consistent with the idea that returns may result from exploiting capital market inefficiencies in small equities. Because small-cap stocks are not otherwise monitored by major institutions, their market prices may deviate substantially from their appropriate values based on careful analysis of all publicly available information.

Another potential source of returns is factor-based inefficiency, in which superior fundamental analysis may identify equity return factors (e.g., size) that differentiate groups of stocks into higher-performing firms that should be held long (e.g., small stocks) and lower-performing firms that should be sold short (e.g., large stocks).

The seminal research of Fama and French (1993, 1995, 2008) indicated that realized U.S. equity returns were driven by factors related to the size of each firm and the extent to which the firm appeared to offer value versus growth. Furthermore, average and perhaps even expected equity returns appear to be related to these factors. Specifically, small stocks seem to outperform large stocks, and value stocks seem to outperform growth stocks over long periods of time. Accordingly, fundamental equity hedge fund managers may have been earning higher returns through their focus on small-capitalization value stocks. These higher returns would be driven by a net long bias, along with the high returns attributable to the size and value factors (Beer 2015).

Numerous questions remain unresolved. Were the higher observed returns of small stocks and value stocks compensation for bearing a higher risk, or were they alpha? Were the higher past returns of small stocks and value stocks a random outcome, or did they occur because small stocks and value stocks offered consistently higher expected returns? Finally, even if the high historical returns reflected superior expected returns in recent decades, will small stocks and value stocks offer higher expected risk-adjusted returns in future decades? With regard to the last question, if capital markets were informationally inefficient in the past with regard to size and growth/value, is there reason to believe that the same markets will continue to be informationally inefficient and continue to offer superior risk-adjusted returns based on these factors?

A few works (Seddighi and Nian 2004; Ahmed 2013) have examined the equity mispricing within developed and emerging countries. Results show a wider range of valuations over time within emerging countries than within developed countries, which would imply that such theories could be justified, at least superficially, given the initial limited empirical support. Somewhat analogously with large stocks versus small stocks, prices in small emerging markets may have a lower degree of informational efficiency than prices in large developed markets. Smaller emerging markets may offer higher returns as compensation for higher risk, as compensation for lower liquidity, or as an incentive for investors to perform the extra fundamental analysis required to invest in relatively inefficient markets (i.e., a complexity premium).

The potentially higher returns of emerging markets might make a reasonably good case for adding fundamental long/short emerging market managers to the common mix of long/short opportunities. A less obvious motivation would be to avoid the home-country biases affecting allocation decisions. Although it may be easier to have due diligence meetings within one's region/location and in one's mother tongue, the advantage of including emerging market managers may be substantial diversification benefits.

An example of potential information that may be used by a fundamental equity manager is observation of trading by other managers. Knowledge of another manager's holdings may be an important piece of information for an investor to use in evaluating the manager. Such knowledge is, to some degree, readily available for some of the hedge funds operating in the United States pursuant to the Securities and Exchange Commission's mandatory reporting rule, otherwise known as Form 13F. Form 13F filings are required of institutional investment managers having discretion over $100 million in 13F securities. These 13F securities include exchange-traded (NYSE, AMEX) or NASDAQ-quoted stocks, equity options and warrants, shares of closed-end investment companies, and certain convertible debt securities. They do not, however, include shares of open-end investment companies (mutual funds) or short sales. The reporting rule requires managers to disclose the names and positions of all 13F securities held long in their portfolios on a quarterly basis within 45 days of the close of the quarter. To the extent that long/short managers do not change their positions frequently, the information obtained through the 13F reports will be current and relevant.

Seasoned long/short fund-of-funds managers examine 13Fs for top positions before meeting new long/short managers. Recently there has been a proliferation of companies that attempt to capture and commercially present this information to clients. Many websites are available that will provide this information to end users for a fee. Additionally, investment products based on 13F positions of popular hedge fund managers have emerged. However, it should be noted that some funds employ legal techniques in order to sidestep this disclosure requirement.

28.2.8.3 Long/Short Return Attribution The benefits and costs of short selling differ between retail investors and institutional investors, and the particular terms and conditions available to various institutional investors vary based on their size, the markets and securities involved, the selection of a prime broker, and so forth. Retail investors must typically post cash collateral on a short sale and are not able to earn interest on the sales proceeds received when shares are sold short. Institutional investors are typically able to earn interest on collateral or to post securities as collateral. Thus, a long/short equity manager could typically post the long positions as collateral on the short positions, thereby generating earnings on the collateral.

The institutional short seller generally receives a short stock rebate, which is interest earned on the proceeds of the short sale. The rebate is typically based on an index-linked variable interest rate (e.g., the federal funds rate) minus the borrowing costs (e.g., 0.15% or 0.25% per year) that the borrower earns on the proceeds of the short sale. In unusual cases of very low interest rates or special-situation securities in which the demand for borrowing is very high, the borrowing costs can exceed the interest revenue, resulting in a negative rebate or a borrowing cost to the short seller. From the security lender's perspective, the lender can borrow money at the rebate rate—a rate below that of a riskless index—which can presumably be invested at a positive spread relative to the rebate.

The following stylized example represents a simplified analysis of typical benefits and costs.

Total returns for long/short managers can be decomposed and attributed to the following four components from the long positions and four components from the short positions.

Returns/costs from long positions:

1. Price appreciation/depreciation
2. Dividends received
3. Margin interest cost of longs (if leveraged)
4. Interest earned on any excess cash or any cash posted as collateral

Returns/costs from short positions:

1. Price depreciation/appreciation
2. Interest earned on proceeds of short sale
3. Cost of borrowing shares, depending on difficulty to borrow
4. Dividend payments to buyers of borrowed shares

EXHIBIT 28.1 Illustration of Cash Flows from Long and Short Positions

Dollar returns from long position (XYZ):	
1. Price appreciation/depreciation	+$10.00 ($110.00 – $100.00)
2. Dividends received	+$ 2.00
3. Margin interest cost of longs (if leveraged)	$ 0.00
4. Interest earned on cash	$ 0.00
Total dollar change from long position	**+$12.00**
Dollar returns from short position (ABC):	
1. Price depreciation/appreciation	+$ 2.50 ($50.00 – $47.50)
2 & 3. Short rebate	+$ 0.75 (0.015 × $50.00)
4. Dividend payments	−$ 1.00
Total dollar change from short position	**+$ 2.25**
Total dollar return	**+$14.25**

Consider a highly simplified example of a long/short manager with $100 of assets under management and only two positions:

1. Long $100 of Company XYZ, which pays a $2 dividend
2. Short $50 of Company ABC, which pays a $1 dividend

Suppose that the objective is to attribute the total performance of this long/short fund over the specific period of one year. Also, assume that no other trades were executed over that year, that XYZ's share price rises 10% over the year, and that ABC's share price declines 5% over the same year. Assume that the short rebate of 1.5% is composed of the rate of return paid to the fund on the proceeds of the short sale (2% per year) and the cost to borrow the ABC shares (0.50% per year). Finally, assume that the fund can post its long shares as collateral for the short sales. Note that there is no excess cash. Exhibit 28.1 illustrates the cash flows.

In the scenario illustrated in Exhibit 28.1, the investor would have earned a gross return of 14.25%, which is the profit of $14.25 divided by the investor capital of $100. In unlevered, fully invested funds, long positions equal the amount of investor capital. In the example, the long side is fully invested and unlevered, so there is no interest income earned on cash and no margin interest expense.

APPLICATION 28.2.8.3

A long/short manager has $600 of assets under management and only two positions: long $500 of Company X, with $7 total per year in dividends, and short $200 of Company Y, with $2 total per year in dividends. Assume there are no other positions over the next year, that X's share price rises 15% over the year, and that Y's share price rises 5% over the same year. Also assume a short rebate of 1% per year and a return on cash of 1% per year. What is the annual total dollar return?

Answer: $73.00 (see Exhibit 28.2).

EXHIBIT 28.2 Cash Flows

Dollar returns from long position (X):	
1. Price appreciation/depreciation	+\$75.00
2. Dividends received	+\$ 7.00
3. Margin interest cost of longs (if leveraged)	\$ 0.00
4. Interest on cash (\$600 – \$500 = \$100)	+\$ 1.00
Total dollar change from long position	**+\$83.00**
Dollar returns from short position (Y):	
1. Price depreciation/appreciation	–\$10.00
2 & 3. Short rebate	+\$ 2.00
4. Dividend payments	–\$ 2.00
Total dollar change from short position	**–\$10.00**
Total dollar return	**+\$73.00**

The prime broker will normally examine a portfolio's characteristics to determine appropriate terms, with more favorable terms, such as lower borrowing and higher rebate rates, generally offered to larger funds. There have been periods in which prime brokers have cut back on the leverage they provided their hedge fund clients, which required leveraged funds to reduce position sizes during volatile markets.

If a long/short fund is fully invested, is unlevered, earns competitive rebates on securities with minimal borrowing costs, and has long securities with similar dividends to the securities it shorts, the returns from a long/short manager should typically be dominated by the price appreciation/depreciation corresponding to the underlying long and short positions, directly reflecting the manager's stock selection skills. Managers can also add value through market timing, whereby net long positions are larger during rising markets and smaller—or even net short—during falling markets.

28.2.9 Fundamental Long/Short Equity Manager's Investment Process

This section reviews a representative fundamental equity long/short manager's investment process. The four procedures are: (1) Understand the business of the company, (2) study the company's management, (3) read and digest the financials, and (4) create and defend the valuation of the company.

1. The Business of the Company

An analysis of the business side of a company takes into consideration two important things: the company's ability to generate stable above-average returns and whether the company has a defensible competitive position. Other factors that would greatly affect the business's performance are its capital requirements and its cash flow profile. Does the company generate the majority of its sales in the holiday season, or do the sales occur consistently across the year? How important is the product or service—is it a need to have or a nice to have? Such questions would help determine the consistencies of the driver of cash flows generated by the business.

Another vital factor in assessing a company is to understand how to measure its base unit. That is, what are the economics of the base unit of the business? Managers must understand exactly how the industry to which this company belongs measures a base unit.

For example, common metrics in the airline industry include available seat miles (ASMs). The ASM unit of measurement refers to one airplane's seat, available for sale, which is assumed to be flown one mile whether occupied or not occupied. In the retail industry, many times the metric of same-store sales is employed. Additionally, in the web-based advertising business, click-through rate (CTR) is an important measurement unit. CTR is the measure of the number of users who clicked a link divided by the total number of impressions (the number of times the ad was served).

Additionally, feedback from suppliers, customers, and competitors concerning a business becomes an extremely vital source of information. When assessing the business's competitive position, this analysis should focus on identifying the company's competitors and comparing the company's performance against how its competitors performed. Additional factors that need to be examined include the company's relative cost position, its market share and product quality, and the talent of its employees.

One of the best examples of this is how Apple's iPhone dominated BlackBerry's devices in the mobile professional communication marketplace. Consider the following hypothetical situation: All the professional users live on an island, and there is a fixed number of users. Each user can use only one mobile device. There are only two mobile products available: iPhone and BlackBerry. Initially, the inhabitants all have BlackBerry devices. As the iPhone is introduced, some early adopters choose to turn in their BlackBerrys and switch to iPhones. The rate at which this switching occurs reveals significant information about the underlying business positions of Apple versus BlackBerry.

Clearly, Apple and BlackBerry have many different business lines beyond their professional mobile communication devices, but this example demonstrates the basic idea of relative positioning, market share, and product quality.

2. The Company's Management

Ideally, the best management is made up of individuals who are highly intelligent, have bulletproof integrity, and sincerely care about the success of the company. Management's compensation and the nature of their equity interest are important factors in determining how much members care about the company. (Their equity interest can be in the form of outright ownership, options, profit sharing, and so on.) When studying the management of a company, it's also important to examine what, if any, checks and balances exist between key members of management.

The company's culture is more difficult to ascertain. One means of understanding the culture is through a careful reading of the company's compliance manual; its code of ethics; and, most important, its policies and procedures, which primarily govern how day-to-day concerns are addressed as they arise at the company. Moreover, studying the CEO and officers of the firm will give a good sense of the company's culture.

Furthermore, a company's culture can also be derived by a study of its employees: Who are they, and what motivates them to work for the company's growth and success? Compare employee compensation levels to those of the company's competitor. Observe the energy level at the company and whether employees come in early and work late. Even go so far as to take into account what their offices look like. Does management micromanage employees? Is there a large turnover in staff? What do ex-employees say about their experiences?

The manner in which management runs the company also sheds important light on the health of the company. Decisions that are made in a centralized rather than decentralized fashion show that management lacks the ability to delegate decisions. When management encourages open communication and feedback, employees typically have more input into business processes. Also important is how the company communicates information pertinent to the business. What information drives the business, and how and when is it received? For example, the speed at which earnings are reported can signal a company that abides by tight standards as opposed to a company that continues to delay such information releases.

Are the company's decisions driven by data or by the whim of the founder/CEO? Is this company a true disrupter or a replicator of the strategies of other firms? How is a company impacted if it has a visionary founder, such as Steve Jobs?

3. Financials

The simplest approach to valuing a company is by looking at its balance sheet and calculating the book value of its equity. The company's book value of equity is its total assets less its total debt. A company's assets are things that the company owns or things for which it has prepaid. Examples of assets are cash, accounts receivable, land, and buildings. A company's debts are its obligations or liabilities owed to creditors. Examples of liabilities are accounts payable, salaries payable, customer deposits, unearned revenues, and income taxes payable.

The **DuPont model**, which is used to calculate ROE (return on equity) based on a company's gross value rather than its net value, can be calculated as follows:

$$\text{ROE} = \text{Profit Margin} \times \text{Asset Turnover} \times \text{Leverage}, \text{ or}$$

$$\text{ROE} = (\text{Net Income/Revenues}) \times (\text{Revenues/Assets}) \times (\text{Assets/Book Equity})$$

The DuPont model provides an organized and uniform way to examine trends over time in the ROE as well as to attribute the performance to the components of ROE.

The following characteristics on the balance sheet warrant special concern because they are critical in showing strengths or weaknesses in a company: capital structure, fixed charges, coverage ratios, changes in working capital, any off-balance-sheet financing, and pension fund issues.

Information about a company's financial condition is also found in its accounting policies and procedures, namely: How does the company recognize revenues? Which depreciation method does the company employ? What are the methods used in recognizing prepaid expenses, asset write-downs, and so on, compared with those of similar companies? Unconventional account practices raise a red flag about a possible attempt to conceal underlying financial problems.

Two other critical areas of a company's financials that need to be taken into account are the manner in which the company spends its cash, and whether the returns generated on new assets are equal to or better than returns on existing assets and the firm's cost of capital.

4. VALUATION

Arriving at a value for a company is no small undertaking for fundamental equity managers because of the breadth and depth of information that may impact a company's value. Managers begin with how the valuation of a company stacks up relative to its peers and the broad market on an earnings, EBITDA, acquisition, or sum-of-the-parts (asset value) basis. From there, the manager also examines the discounted cash flow valuation model. Because future profitability is the bottom line, the way the market values the company 18 months hence, and how that valuation looks relative to historical norms, is part of the deep analysis.

EARNINGS APPROACH. Another way to value a company is to compare its earnings with those of similar companies. Earnings are also referred to as net income, since they are defined as the money remaining after all the company's bills have been paid. A company's earnings are represented as earnings per share (EPS), which indicates how much net income the company generated for one share of its stock. Though the EPS ratio allows for direct comparison with other companies, this ratio on its own does not give much insight as to the profitability of the company because it doesn't consider the price of each share of the company, which is why the price-to-earnings (P/E) ratio is preferred. The P/E ratio takes a company's stock price and divides it by its EPS over the past four quarters. For example, take two companies:

COMPANY A

$5,000,000 net income

500,000 shares

$10 in earnings per share

$200 share price = 20 P/E

COMPANY B

$10,000,000 net income

500,000 shares

$20 in earnings per share

$200 share price = 10 P/E

In general, most analysts prefer a low-P/E stock, which makes Company B a more attractive investment opportunity than Company A. But there are serious shortcomings to the use of the P/E ratio in isolation. Due to investment technology increasing the efficiency of the market, it is rare to find a company with a low P/E ratio that has gone undetected by other bargain hunters. This means that if a company has a low P/E ratio, it is most likely not an underpriced stock but a stock with possible risks, such as high debt levels or slow or negative growth in sales or net income. Remember that a stock's P/E ratio is backward looking, as it takes into account past earnings. In order to determine if a stock's low P/E ratio is not a sign of future trouble, a valuation tool that measures expected future profitability is necessary.

28.3 GLOBAL MACRO

Global macro managers have been around for more than 30 years, but their golden era culminated in the early 1990s, when their strategy represented more than half of worldwide hedge fund assets. This was the time when legendary managers such as George Soros (Quantum Fund) and Julian Robertson (Tiger Management) were running multibillion-dollar funds involved, primarily, in leveraged directional trades. In 2015, discretionary macro funds represented less than 10%, or nearly $300 billion, of the hedge fund industry's assets under management. While macro assets have grown over time, they have grown at a slower rate than other hedge fund strategies. Investors' appetite for global macro funds fell in mid-2000, when high-profile operators such as Soros and Robertson shut down their macro funds and retired after posting disappointing performance numbers.[5] Other fund managers returned capital to investors because they were losing their edge in large liquid markets yet were too big to operate in illiquid markets. More important, by 2005, the lack of volatility across global markets had made it difficult for global macro funds to make money. Last, the number of liquid bond and currency markets (and hence opportunities) was drastically reduced with the introduction of the euro.

After a long dry spell, the global macro strategy came back in favor in the second half of 2007 and in 2008, when global market volatility soared and the fear of a U.S.-led global recession became a real threat. Short dollar positions and long commodity trades became clear winners for global macro funds in 2007, with the opposite positions profiting in the second half of 2008. But the global macro funds of today are very different from those of prior generations. Leverage is still employed, but the focus is more on the consistency of returns and effective risk management.

28.3.1 Introduction to Macro Strategies

Global macro hedge funds have the broadest mandate of any of the major fund strategies. Their mandate often has no limitations in terms of types of instruments, asset classes, markets, and geographies. They can dynamically allocate capital to the asset class, sector, or region in which they think the best opportunities currently lie—hence the term *global*. The second term, *macro*, reflects the fact that these managers apply macroeconomic views to global markets. Instead of analyzing microeconomic events affecting companies or assets, they view the world from a top-down perspective. Their goal is to anticipate global macroeconomic changes and themes, detect trends and inflection points, and profit by investing in financial instruments whose prices are likely to be impacted most directly. They can go long or short, be concentrated or diversified, and use leverage or not. While some funds trade single stocks in anticipation of macro themes, most funds concentrate trading in forwards, futures, and swaps on macro markets, including commodities, currencies, equities, and interest rates.

There are probably as many approaches to the strategy as there are global macro hedge fund managers, but they share a common desire to identify and exploit markets in severe disequilibrium. It is only when prices are perceived as being more than one or two standard deviations away from the fair value that macro traders deem that the market presents a compelling opportunity. Ideally, macro managers will be able to find trades with an asymmetric risk-reward profile, offering minimal potential losses and large potential gains.

28.3.2 Discretionary versus Systematic

Discretionary global macro managers usually perform intensive fundamental research. They continuously analyze information from varied sources, such as central bank publications, survey data, confidence indicators, asset flows, liquidity measures, forecasting agencies, political commentators, and personal contacts. They look for markets that do not match the macroeconomic realities, identifying situations that are unsustainable or asset classes that are likely to follow predictable trends. They spend hours forming their views on likely market scenarios while assessing the probabilities of alternative scenarios. Once interesting risk-reward opportunities have been identified, global macro managers determine appropriate entry points, often by applying traditional technical analysis. The overall result is completely discretionary and highly dependent on the particular skills of the manager.

By contrast, systematic global macro managers apply a highly structured, disciplined, and repeatable investment process. They replace subjective macroeconomic analysis with a systematic way of looking at economic data, and rely on mathematical models to evaluate markets, detect trading opportunities, generate signals, and establish entry and exit points. Many of them seek to identify specific fundamental data and key economic drivers that explain the long-term behavior of various markets, and combine these elements with assessments of the current economic conditions and market forecasts. Others focus on identifying directional trends in markets that can be traded in large volumes and where capital is capable of moving quickly. Their systems can be based on moving averages, breakout systems, pattern recognition, or any combination of these. Some investors like to call them global trend followers, as they are very close in spirit to trend-following commodity trading advisers (CTAs), particularly when they focus on longer-term trends.

The reality is that CTAs and some global macro funds share a fundamental source of returns—namely, long-term secular shifts in capital flows. Generally, both tend to participate in large trends in the major equity, fixed-income, and foreign exchange markets and, to a lesser extent, the energy, agriculture, and metals markets. But they participate in these trends in four different ways:

1. CTAs are reactive, whereas global macro managers are often anticipatory. They therefore often overlap in the middle part of a well-established trend, but their entry and exit points are fundamentally different.
2. CTAs are purely price based in their analysis and follow their systematic models regardless of fundamentals. Global macro managers prefer to look at the big picture and stand aside when market fundamentals do not appear to explain a trend properly.
3. CTAs tend to exhibit the same views on markets at the same time as one another because their inputs and systems are similar. By contrast, global macro managers can be quite different from one another in how they evaluate data and make trading decisions.
4. CTAs are generally momentum (technically driven) traders, whereas their global macro counterparts focus on such fundamentals as inventories in commodity markets or interest rate differentials in currency markets.

These disparities in approach help explain the disparity of returns and return volatility.

28.3.3 Global Macro Schools of Thought

Alternatively, one may want to distinguish the sources of returns that global macro funds are trying to tap. In this case, according to Ahl (2001), there are essentially three possible schools of thought to be considered:

Feedback-based global macro managers assume that markets are rational most of the time but that there can exist periods of severe irrationality. Such periods can arise either because people have made money too easily and become complacent or because they have lost money too quickly and become stressed or distressed. As a result, feedback-based global macro managers attempt to read the financial market's psychology, sell into bursting bubbles, and buy into post-crash recoveries.

Information-based global macro managers rely primarily on collecting micro-level information to better understand the global macro picture. Their hypothesis is that an information gap is created by the delay in the release of official macroeconomic statistics. This gap then opens the door for pricing inefficiencies, which will persist until the macro information has been disseminated into the public domain.

Model-based global macro managers rely primarily on financial models and economic theories to analyze market movements, detect policy mistakes of central banks and governments, or extract implied market expectations and compare them to sensible estimates. As discussed by Safvenblad (2003), these are five examples of trades or models commonly used:

1. Carry trades (i.e., using interest rate differentials as indicators of positive carry positions) involve holding long-maturity bonds against shorter instruments, or holding high-yielding currencies against low-yielding ones.
2. Yield curve–relative value trades involve identifying the undervalued or overvalued part of the yield curve or trading the slope of one yield curve against the slope of another.
3. Purchasing power parity models are often used by global macro funds to assess the relative value of currencies.
4. Valuation models used in trading equity markets, such as the dividend discount model, are usually applied bottom-up at each company level; expected returns for each company are then aggregated and weighted to derive a market-level expected return at a country level.
5. Option pricing models provide the market's implied views about the future volatility of some underlying asset.

28.3.4 Multistrategy Global Macro Funds

Investing large sums of money is usually not an issue for global macro managers, given their flexibility and the depth and liquidity of the markets they trade in; however, the reality is that, past a certain fund size, it becomes prudent to add more traders and strategies. Thus, it should come as no surprise that the larger global macro–oriented hedge funds have migrated to a multistrategy model, which in turn has increased their correlations with funds of hedge funds. In such a case, identifying precisely which school of thought a manager belongs to may be difficult. For this reason, this chapter approaches the global macro universe from a trading perspective rather than from a classification perspective.

28.3.5 Directional Currency Trades

Directional currency trading is driven by relative values of currencies. Accordingly, our analysis of this topic begins with a discussion of the linkage between interest rates and the inflation rate of the currency in which the interest rate is expressed.

THE FISHER EFFECT. Prior to delving deeply into a discussion on global macro strategies, and currency trading in particular, we provide a brief review of relevant theoretical concepts on inflation, interest rates, and currency movements. The Fisher effect (Fisher 1930) is one of the main pillars or building blocks that must be well understood to discern the dynamics of the fixed-income and currency markets. In particular, the Fisher effect—also known as the Fisher equation—states that the nominal interest rate (n) is approximately equal to the real interest rate (r) plus the expected inflation rate (i). The Fisher equation can be expressed as

$$n \approx r + i \tag{28.8}$$

or, more accurately,

$$(1 + n) = (1 + r)(1 + i) \tag{28.9}$$

Fama (1975) empirically tested the Fisher equation using historical Treasury bill rates from 1953 to 1971. Over this period, he found strong empirical evidence documenting that the Fisher equation held.

Many assume that real interest rates are constant, so any changes over time in expected inflation are captured by changes in the nominal rate. For example, if the current Treasury bill rate is 2% and the real interest rate is constant at 1%, then the expected inflation would be 1% over this given period. Furthermore, if the Treasury bill rate increased to 4% over the next year and the real interest rate remained constant at 1%, then it is deduced that expected inflation would have increased to 3%. Over time, expected inflation increased from 1% to 3% in our example, and the nominal interest rate increased from 2% to 4%.

CURRENCY FLUCTUATIONS. The basic intuition in currency fluctuations can be viewed as employing the Fisher equation. Let's assume that we have two countries, country C and country D, with their own currencies. If we assume that both countries have zero inflation, then the difference in nominal interest rates will be equal to the differences in real interest rates. Let's further assume that country C has a real interest rate of 2%, and country D has a real interest rate of 1%. If we further assume that capital can flow effortlessly from country to country, we would see capital flowing from country D's low real rate of 1% to country C's high real rate of 2%. Since the process of purchasing goods in country C requires transactions in country C's currency units—let's call them \$C—then the demand for \$C would increase and the demand for \$D would decrease. The currency exchange rate of \$C per 1 unit of \$D would increase. \$C would strengthen relative to \$D, so people in country D would keep converting their \$D to \$C to buy goods in country C, as such exchange rates would settle to reflect this relatively higher demand for \$C.

Now let's assume real rates are constant for both country C and country D. We could go further and say they are 1% for both countries. Now suppose the nominal interest rate in country C is 10%, so 9% is due to expected inflation. With a nominal interest rate of 5% in country D, 4% is due to expected inflation. Since inflation reflects a decrease in purchasing power over time, country C's future is bleak relative to country D's future. So, people in country C would start exchanging \$C for \$D to invest in country D's goods. Thus, we would see country C's currency depreciate versus country D's currency. In this case, the reverse of the previous scenario would happen. The main point is that the composition of the nominal interest rate should drive the currency differences. If the reason for the difference in nominal rates is purely because of inflation, the higher-inflation currency would be unattractive relative to the lower-inflation currency. Moreover, if the reason for the difference in nominal rates is solely due to real rates, then the reverse would occur. In the real world, it's much more complicated, since we cannot observe real interest rates. Most of the time both real interest rates and inflation are changing over time. Generally, real rates are changing slightly over time, and inflation may be changing substantially if a country's monetary policies are in play. The impact of monetary policies on inflation, although a fascinating subject, is a hotly debated topic and beyond the scope of this chapter.

Background on currency trading. One would be remiss, in discussing the global macro strategy, to neglect to mention George Soros and the Exchange Rate Mechanism (ERM) crisis of 1992–1993 (see Exhibit 28.2). This was likely the first time the general public became aware of the existence of global macro funds and their actions. This case study explains the discretionary, fundamental, information-based style of macro trading. In this case, the information used focuses on central bank policies and the level of currency reserves held by each country.

The European Monetary System (EMS) was formed in 1979 by several European countries to coordinate their monetary and exchange rate policies. Among

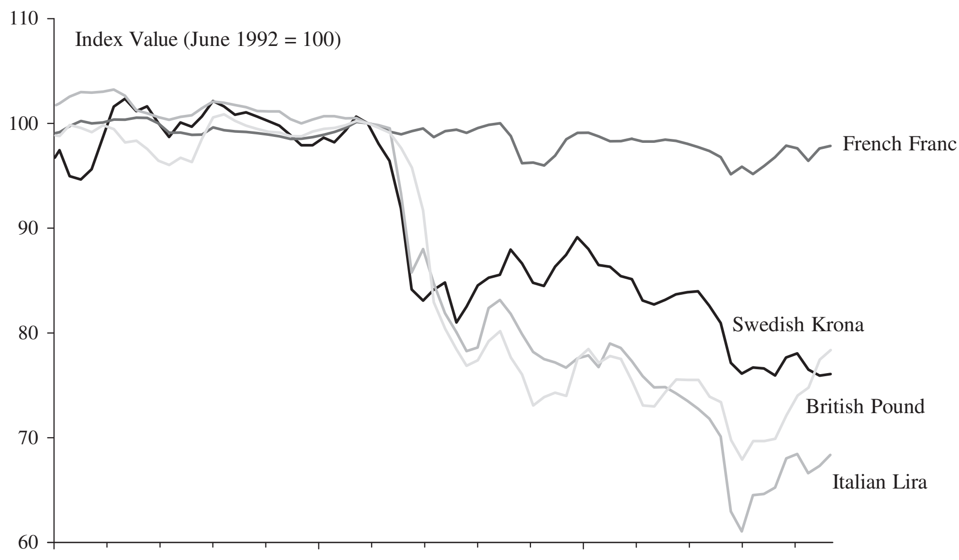


EXHIBIT 28.2 Evolution of Exchange Rates around the 1992 Crisis

other things, the EMS created a fictitious accounting unit called the European Currency Unit (ECU), using guidelines established by the ERM agreement. The ERM was essentially a managed floating exchange rate system, in which the currencies of participating countries were allowed to fluctuate—within prespecified bands—around a reference point (±2.25% for most countries; ±6% for Italy, Spain, Portugal, and the United Kingdom). Central banks were charged with taking appropriate measures, whenever needed, to keep the exchange rate within these bands. Since the ECU was fictitious, in practice the unofficial reserve currency, the German mark, turned out to be the most stable currency of the group. That is, the bands were effectively maintained with respect to the German mark, and central banks typically intervened by selling or buying marks against their respective domestic currencies. In a sense, the role of Germany within the ERM was similar to that of the United States within the Bretton Woods system. Since there were only 10 direct exchange rates among the 11 member countries, Germany was free to set its monetary policy, often forcing other countries to follow its lead, even when doing so ran counter to domestic interests. In practice, the countries had effectively lost all control over their own monetary policies. When an exchange rate approached one extreme of the preset bands, the respective central bank was forced to intervene, using the country's reserves to maintain the band.

Since its creation, the EMS had achieved its main objectives to a large extent. In 1992, however, the policy interests of Germany and the rest of Europe began diverging dramatically. German reunification created strong domestic growth, which drove German interest rates higher and ultimately pushed other European currencies to the bottom of their respective bands. By contrast, the United Kingdom was in its worst recession since the end of World War II, with unemployment rates well in excess of 10%. In the absence of the ERM agreement, the United Kingdom would have resorted to an expansionary monetary policy or a devaluation to get out of the slump. Unfortunately, its actions were constrained by the fixed exchange rate system.

While most market participants thought that it was impossible to build up enough pressure to force the British authorities to abandon the bands, a few speculators, led by George Soros and his Quantum Fund, decided to launch a speculative attack. In the summer of 1992, they sold a huge amount of British pounds in exchange for other currencies. The Bank of England rushed to defend the band through intervention but rapidly depleted its foreign currency reserve holdings in the process. On September 3, to replenish these lost reserves, the Bank of England was forced to borrow 10 billion ECUs (i.e., approximately $14.3 billion) on the international market. At the same time, speculators began attacking the Italian lira, forcing the Bank of Italy to raise its discount rate from 12% to 15%. The Italian central bank may have raised interest rates to attract new capital flows and to increase the cost of holding short positions in the lira.

Speculation against the pound culminated on September 16, subsequently referred to as Black Wednesday.[6] Exacerbated by uncertainties over a French referendum on European construction, massive speculative flows continued to disrupt the functioning of the Exchange Rate Mechanism. The Bank of England responded by raising its base lending rate from 10% to 12%, then to 15% the next day, but both announcements had little impact on the intense speculative pressure. Massive interventions in the foreign exchange markets proved futile as well. Ultimately, Germany agreed to lower its discount rate by 50 basis points and its Lombard rate by 25 basis points, but by then it was too late. Although the Conservative government had repeatedly vowed otherwise, the Bank of England was forced to suspend the participation of the pound in the ERM. Italy soon followed, causing a rapid and massive depreciation of both currencies. Spain was also pressured to devalue its currency by 5% and impose capital controls. The speculative attacks continued well into 1993, when the bands of several ERM currencies were widened temporarily to 15%, which meant they were effectively floating currencies. For George Soros and his $10 billion Quantum Fund, the result was a profit of more than $1 billion in this directional bet against the British pound.

More than 20 years later, there are still opportunities to trade currencies with a view toward changing governmental policy. Many investors have long positions in the Hong Kong dollar (HKD), anticipating that the HKD is undervalued and likely to break its peg to the U.S. dollar (USD) and revalue the currency from its official trading range of between 7.75 and 7.85 HKD to the USD. Other macro traders are attracted to the currencies of the Gulf Cooperation Council (GCC) countries (United Arab Emirates, Saudi Arabia, Qatar, Oman, and Bahrain). The Saudi riyal has long been pegged at an official rate of 3.75 to the USD. If energy prices change substantially and capital flows into or out of the region, these countries may feel pressure to move to floating currencies, likely resulting in a substantial change in valuation versus the USD.

28.3.6 The Case of Emerging Markets

The 1997 Asian currency crisis brought to the foreground concerns about global macro funds and their possible role in exacerbating financial market volatility and disrupting emerging markets. Some Asian government officials explicitly accused hedge funds of attacking their currencies and causing their downfall. Specifically, Malaysia's prime minister at the time, Mahathir Mohamad, argued that by accumulating very large and concentrated short speculative positions (referred to as "big elephants in small ponds"), hedge funds had destabilized the foreign exchange, money, and equity markets of Thailand, Malaysia, Indonesia, and the Philippines. Several governments also raised concerns about aggressive and manipulative tactics used by some global macro hedge funds, which might have compromised market integrity and interfered with a normal price-discovery process; see Brown, Goetzmann, and Park (2000) for further discussion.

While the trading activity of hedge funds may have sped up the devaluation process, there was no doubt that many of these currencies were fundamentally weak. Black (2004) explains how the supply of and demand for a currency can be monitored using the current account and capital account balances as well as the level of official reserves. For freely floating currencies, a country with net capital inflows would likely see an appreciating currency, whereas a country with net capital outflows would likely experience depreciation. The **current account** measures the balance of trade in goods and services, whereas the **capital account** measures the net flow of financial transactions. For most countries, there will be a **balance of payments**, as the capital account and current account will have offsetting flows, such as when a current account deficit is offset by a capital account surplus. In a fixed-rate or pegged currency, net capital outflows cause the central bank to purchase the domestic currency. Once the central bank has exhausted its official reserves, there are no assets left to support the currency at the pegged price, and the currency must be devalued to a point where the central bank is no longer required to purchase the currency to support its value. Many Asian currencies in 1997 and 1998 were pegged at levels above fair value, and the devaluations that were profitable to hedge funds were ultimately the result of both trade imbalances and investment outflows.

The International Monetary Fund (IMF) responded to the charges of market manipulation by examining the role of hedge funds in the Asian currency crisis. In the resulting study, Eichengreen et al. (1998) find no evidence that hedge funds played a major role in the events leading up to the Asian crisis, but many governments contested these conclusions as not fully coming to grips with the role played by hedge funds.

Political sensitivities aside, it can be difficult to fully assess the role and impact of hedge funds in emerging markets. First, many hedge funds operate through over-the-counter foreign exchange and money markets, which are very opaque. Second, hedge funds themselves are opaque, as they are not subject to mandatory reporting requirements outside of some long equity positions. As a result, any assessment of the impact that hedge funds may have had in Asia in the late 1990s necessitates access to private information (e.g., trading and position data) or market intelligence (such as off-the-record meetings with various prime brokers and traders), or requires assumptions to be made in order to build an econometric model.

Brown, Goetzmann, and Park (2000) investigated the changing positions of the top 10 global hedge funds vis-à-vis the Malaysian ringgit. They conclude that there were periods when hedge funds had large long and short currency exposures, but that changes in these positions had no relationship to exchange rate movements. In their opinion, there was no evidence to support a claim that hedge funds in general—or any hedge fund in particular—led the charge against Asian currencies.

Using data on the returns of 27 large hedge funds during the Asian currency crisis, Fung, Hsieh, and Tsatsaronis (2000) attempted to infer the currency positions held by the funds. They compare those inferred positions with the total capital flows for the Asian countries based on balance-of-payments accounts, and conclude that aggregate hedge fund positions were too small to have caused the collapse of the Asian currencies.

Last, the Financial Stability Board (2000) also attempted to assess the impact of hedge funds in Asia during the crisis. It expresses some concerns about the apparent large size and concentration of macro hedge fund positions and their implications for market dynamics, especially during unsettled market conditions. It provides evidence on the aggressive activities of some hedge funds “talking their books” (making statements supporting investment positions, spreading rumors, aggressively trading at illiquid times), which could be seen as efforts to move markets. But the report also points to the downside risk of using size to move a market. Large players, such as global macro hedge funds, face strong incentives to tailor the size of a position to market size and liquidity in order to avoid moving prices too far when unwinding positions.

28.3.7 Four Models for Currency Trading

Pojarliev and Levich (2008) describe four models for currency trading: carry, trend following and momentum, value, and volatility. Rather than assuming that all profits from currency trading are alpha, their research shows that manager skill for currency and global macro managers should be measured as the returns in excess of those earned from beta exposure to each of these four factors. From 1990 to 2006, Pojarliev and Levich calculated total returns to the four currency strategies combined at 62 basis points per month, with the risk-free rate accounting for 37 basis points and the remaining 25 basis points representing the excess return of the strategies. The 25 basis points of excess return were concentrated in the years prior to 2000. In fact, after 2000, the excess return to the four strategies declined to just 8 basis points per month. The carry and trend-following factors were profitable, but the value and volatility strategies had returns near zero over the full period. With a monthly standard deviation of excess returns of 3.04%, all four strategies combined had a very low information ratio. The trend-following and volatility return factors had the highest volatility, whereas the carry and value factors had much lower volatility. Their regression models of the returns of a basket of currency managers on the four strategy factors had an *r*-squared of 0.66, meaning that the majority of returns to currency funds were attributable to these four factors. In their analysis of individual manager returns, approximately one-quarter of the individual managers had the majority of their returns explained by the four factors. Further, approximately only one-quarter of the currency managers had a positive and statistically significant alpha after adjusting for the four-factor returns and fees.

Both carry and trend-following/momentum strategies appear to be relatively simple to implement, and yet, surprisingly, they have been shown to be profitable on average. Naturally, this raises the question about the sources of the profitability of these two strategies. Four common answers to this question are:

1. The profitability of these two strategies represents fair compensation for the systemic risk to which they are exposed.
2. The profitability is not real and will disappear when transaction costs and market impacts are accounted for.
3. The profitability is highly unpredictable, which limits arbitrage on the part of traders.
4. The profitability results from trading against central banks that intervene in currency markets.

Of these four answers, the first does not provide a satisfactory explanation, as academics and practitioners have not been able to identify the specific sources of systematic risks associated with these two strategies. That is, exposures to credit risk, equity risk, interest rate risk, and so on do not seem to explain returns to these two strategies.

The second explanation does provide a partial answer as far as returns from momentum strategies are concerned. Research shows that momentum strategies are most profitable when applied to thinly traded currencies. Therefore, the actual profitability of a momentum strategy will be lower because of the transaction costs associated with trading the currencies of smaller economies. These results do not apply to the carry trade, as this strategy has been profitable when applied to the most active liquid markets.

The third explanation applies to both strategies and may help to explain the persistence of the profitability of these two strategies. Historically, both strategies have been profitable, but there have also been periods of significant losses. Both strategies tend to attract large pools of capital, which means their trades could become crowded, leading to large losses when trends reverse. For example, the carry trade involving the Japanese yen was very popular in 2007 and early 2008. However, when these trades began losing money in the first quarter of 2008, the rush by traders to close their positions exacerbated the situation, leading to large losses.

The fourth explanation does explain some profitable trades (e.g., the famous trade by George Soros against the Bank of England, which reportedly netted the Quantum Fund $1 billion in profits). However, central bank interventions do not appear to be widespread enough to explain the persistent profitability of these two strategies.

28.3.8 Carry Models for Currency Trading

A carry trade is a very simple yet popular global macro strategy. In its simplest form, a carry trade in currency markets consists of borrowing in a low-interest-rate currency and lending in a high-interest-rate currency without hedging the exchange rate risk. The goal of such a trade is to capture the interest rate differential, which can be quite substantial. As an illustration, if the funding currency carries an annual interest rate of 0.5% versus 5.25% in the target currency, a trader would borrow in the funding currency, convert the proceeds into the target currency, and buy bonds using the target currency. This would earn the trader a positive carry (or profit) of 4.75% as long as the exchange rate does not change. Carry trading remains profitable as long as the target currency does not depreciate by more than the interest rate differential. This explains why global macro funds often establish carry trades between currencies that display both high interest rate differentials and low exchange-rate volatility.

Currencies such as the Japanese yen and, more recently, the Swiss franc have been popular choices for the borrow side of carry trades due to their low interest rates. For instance, one can borrow yen at rates as low as 0.5%. By contrast, high-yielding currencies—such as the South African rand, the New Zealand dollar, the Brazilian real, or the Icelandic krona—have been attractive currencies in which to lend at high interest rates.

To some extent, the profitability of a carry trade can be reinforced by the flows it generates. For instance, a yen-financed carry trade requires the trader to sell the borrowed yen to convert them into higher-yielding assets in order to earn a profit. As large amounts of yen flow out of Japan into the global trading system, the yen weakens (all else being equal) while the target currency appreciates, thus reinforcing the cycle. However, at some point the carry trades will unwind. That is, traders will sell the assets in the target currency and convert the proceeds into the funding currency in order to pay back their initial loans. This typically occurs when one or more of the following three things happen:

1. The funding currency interest rate rises, thereby increasing borrowing costs.
2. The funding currency appreciates against the target currency.
3. The target currency investment does not yield as much as initially expected. This risk is particularly important in more aggressive carry trades, in which the investments made using the target currency consist of high-yielding assets, such as non-investment-grade corporate bonds or even equities.

When performed on a large scale, the unwinding of carry trades may be brutal and can lead to significant amounts of financial market volatility, especially if many traders exit en masse from the same positions at the same time. Carry trades frequently unwind during times of elevated systemic risk, experiencing losses at the same time as long investments in equity markets and arbitrage-strategy hedge funds. For instance, in the summer of 1998, the Japanese yen, which had been depreciating versus the dollar for three years, started appreciating. Traders rushed to sell their high-yielding assets and their underlying currencies against the yen in order to pay back their yen-denominated loans. Not only did the Japanese currency appreciate very sharply in early October as investors scrambled to buy yen, but prices of high-yielding assets depreciated quickly, forcing the U.S. Federal Reserve to cut the federal funds rate twice (for a total of 50 basis points) in the subsequent month to restore liquidity to the financial markets.

In January 2015, the Swiss franc appreciated 15% relative to the euro. Investors who used the Swiss franc as the low-yielding currency in a carry trade suffered large and sudden losses. Many hedge funds closed out this trade at a large loss rather than choosing to remain in a short franc position.

To understand carry trades, two related models of currency markets must be explained. The first model, referred to as **covered interest rate parity**, is similar to the cost-of-carry model of commodity markets, and it relates the spot and forward exchange rates to differences in short-term interest rates in the two countries.

To explore this concept in mathematical terms, let S_t(DCU/FCU) denote the spot value of country D's currency unit (DCU) in terms of units of country D's currency per unit of country F's currency (FCU). Also, let r_{DCU} and r_{FCU} denote annualized short-term riskless interest rates on instruments denominated in DCUs and FCUs, respectively. Finally, let F_t denote the current forward rate on a one-year forward exchange rate contract. According to covered interest rate parity, the covered (hedged) return from investing in the FCU-denominated instrument should be equal to the rate of return on the DCU-denominated instrument. That is,

$$(1 + r_{FCU}) \times \frac{F_t}{S_t} = (1 + r_{DCU}) \tag{28.10}$$

The left-hand side of Equation 28.10 is the hedged return from investing in the FCU instrument, while the right-hand side represents the return from investing in the DCU instrument.

APPLICATION 28.3.8a

Suppose annual short-term interest rates in the United States and Japan are 4% and 1%, respectively (i.e., r_{FCU} = 4% and r_{DCU} = 1%). In addition, the current spot rate and the one-year forward rate for the yen (JPY) versus USD are 120 and 116.54, respectively (i.e., 120 JPY = 1 USD in the spot market, and the forward market is quoting 116.54 JPY = 1 USD). Does covered interest rate parity hold, assuming zero transaction costs?

Because the exchange rates are expressed as yen per dollar, the yen is the DCU, and the U.S. dollar is the FCU. Thus, (1 + 0.04) (116.54/120) = (1 + 0.01). It can be verified that this equality holds, and covered interest parity is satisfied.

If covered interest rate parity does not hold, then traders can take advantage of the situation by investing in the instrument that has the higher hedged rate of return. For example, if the short-term rate in Japan were 1.1%, then traders could generate an arbitrage profit by shorting U.S. bonds and using the proceeds to purchase Japanese bonds, while hedging against currency risks in the forward market. This would generate a one-year riskless arbitrage profit of 0.1%, not including transaction costs. Because potential profits are essentially riskless, there are rarely meaningful deviations from covered interest rate parity.

Uncovered interest rate parity is similar to covered parity, with the transactions left unhedged. Using the same notation, uncovered interest rate parity holds if the following equality holds:

$$(1 + r_{FCU}) \times \frac{E[S_{t+1}]}{S_t} = (1 + r_{DCU}). \tag{28.11}$$

In the previous example, it was stated that there should be no arbitrage profit from borrowing in one currency and investing the proceeds in another currency while hedging against currency risk in the forward market. According to the uncovered interest rate parity model, even if the transaction is not hedged against currency risk, the profit is expected to be zero.

APPLICATION 28.3.8b

Suppose annual short-term interest rates in the United States and Japan are 4% and 1%, respectively. In addition, the current spot rate for the yen per the USD is 120 (i.e., S_t= 120). What is the implied expected future spot rate in one year according to uncovered interest rate parity, assuming zero transaction costs?

According to the uncovered parity model, the income from investing in USD and Japanese bonds should be the same after adjusting for currency changes. The yen return on an investment in U.S. bonds is (1 + 0.04) × (S_{t+1}/120).

The yen return on an investment in Japanese bonds is 1%. These two returns are expected to be equal to each other according to uncovered interest rate parity: (1 + 0.04) × [$E(S_{t+1})$/120] = (1 + 0.01). Thus, the expected spot rate at time period t + 1, $E(S_{t+1})$ = [(1 + 0.01)/(1 + 0.04)] × 120 = 116.54.

That is, the yen/USD spot rate is expected to be 116.54. The implication is that the currency of the country with the higher interest rate—USD in this case—should decline versus the currency of the country with the lower interest rate. At first this appears to be counterintuitive, as one would think the currency of the country with the higher interest rate would be more attractive and would therefore appreciate. If that were the case, the return from investing in the currency of the high-interest-rate country would be highly profitable, because the investor would enjoy not only the benefits of the higher interest rate but also the benefits of an appreciating currency.

The carry trade is, in fact, based on the belief that uncovered interest rate parity does not hold on average, and that currency of the high-interest-rate country does not depreciate enough to completely offset the benefits of the higher interest rate. In particular, a carry trade is implemented using the following rule: Borrow in terms of currencies with lower interest rates, and invest the proceeds in instruments that are denominated in the currencies with higher interest rates. The trade will be profitable as long as the high-interest-rate currencies do not depreciate enough to offset the differentials between the borrowing rates and the returns on the investment. In this case, gains or losses from the carry trade can be expressed using the following rule:

$$\text{P \& L from Carry Trade} = \begin{cases} (1 + r_{FCU}) \times \dfrac{S_{t+1}}{S_t} - (1 + r_{DCU}) & \text{If } r_{FCU} > r_{DCU} \\ (1 + r_{DCU}) - (1 + r_{FCU}) \times \dfrac{S_{t+1}}{S_t} & \text{If } r_{DCU} > r_{FCU} \end{cases} \tag{28.12}$$

APPLICATION 28.3.8c

Suppose short-term rates in Brazil and the United States are 4% and 2%, respectively. The current value of the Brazilian real (BRL) versus the USD is 0.53. Suppose that after one year the spot value of BRL versus USD is 0.51. Has the carry trade been profitable, assuming zero transaction costs?

First, because the short-term rate is higher in Brazil than in the United States, the carry trade requires an investor to borrow short-term in the United States at 2% per year and invest the proceeds in BRL-denominated instruments earning 4% per year. Suppose the investor borrows $1 million. After conversion into BRL, the investor invests (1/0.53) million in BRL-denominated instruments. After one year, this investment will grow to (1/0.53) × (1 + 0.04) million BRL, and after conversion into USD, it will be (1/0.53) × (1 + 0.04) million × 0.51 = 1.0007 million USD.

This sum is less than what the investor will need to repay the loan acquired in the United States: USD 1.02 million. In this case, the decline in BRL has more than offset the gain from investing in higher-yielding instruments denominated in BRL. This carry trade has generated a loss.

The carry trade can also be implemented using forward and futures contracts. Given the results reported by covered interest rate parity, it can be seen that the return from investing in higher-interest-rate currencies would be similar to taking a long position in a futures contract on the higher-interest-rate currency while taking a short position in the futures contract on the lower-interest-rate currency. In other words, take a long position in FCU or a short position in DCU if $F_t < S_t$, or a short position in FCU or a long position in DCU if $F_t > S_t$.

The long strategy will be profitable if the future spot rate, S_{t+1}, turns out to be less than the current forward rate, F_t Similarly, the short strategy will be profitable if the future spot rate, S_{t+1}, turns out to be greater than the current forward rate, F_t When forward or futures contracts are used to implement the carry trade, gains or losses can be calculated using the following rule:

$$\text{P \& L from Carry Trade} = \begin{cases} F_t - S_{t+1} & \text{If } F_t > S_t \\ S_{t+1} - F_t & \text{If } F_t < S_t \end{cases} \tag{28.13}$$

Empirical academic and industry studies have shown that, on average, carry trades have been a profitable strategy. For example, a study by Burnside, Eichenbaum, and Rebelo (2011) shows that a strategy that had invested in the currencies of high-interest-rate countries between 1976 and 2009 would have earned an average return of 4.6% per year. This may not appear to be a significant figure in economic terms, but when it is noted that the strategy could be implemented using futures contracts, where capital commitments are relatively small, this becomes economically significant. Further, the Sharpe ratio of this strategy is more than twice as large as the Sharpe ratio of a buy-and-hold strategy using the S&P 500 Index.

28.3.9 Trend-Following and Momentum Models for Currency Trading

A momentum strategy is concerned with the past performance of currencies rather than with interest rate differentials. To implement this strategy, an investor would examine the relative performance of a set of currencies, and then take long positions in those currencies whose values have appreciated in the recent past and take short positions in those currencies whose values have depreciated in the recent past. Using the notation employed earlier, the gains and losses from a currency momentum strategy can be calculated using the following rule:

$$\text{P\& L from Momentum Trade} = \begin{cases} S_{t+1} - S_t & \text{If } S_t > S_{t-1} \\ S_t - S_{t+1} & \text{If } S_t < S_{t-1} \end{cases} \tag{28.14}$$

This means that if the FCU increased in value during the recent past, one should take a long position in the FCU, and thus the strategy would show a profit if the trend continues and $S_{t+1} > S_t$ In contrast, the long strategy would show a loss if the trend reverses and the FCU declines in the next period—that is, $S_{t+1} < S_t$.

Empirical academic and industry studies have shown that, on average, currency momentum trades have been profitable. For example, the same study by Burnside, Eichenbaum, and Rebelo (2011) shows that a strategy that took long positions in a portfolio of appreciating currencies and short positions in a portfolio of depreciating currencies between 1976 and 2009 would have earned 4.5% per year. The return on a momentum strategy is more volatile than the return from a carry trade strategy and, therefore, has a lower Sharpe ratio of 0.62. Further, studies have shown that a significant portion of profits generated by a momentum strategy comes from currencies of smaller countries, which may suggest that the profits are compensation for the risk of holding these currencies. As determined by Menkhoff, Sarno, and

Schrimpf (2012), transaction costs may significantly reduce the reported profits of the momentum strategy.

28.3.10 Value and Volatility Models for Currency Trading

While momentum models may explain a large portion of currency returns in the short run, value models may be more successful at explaining long-run returns to currencies. In the carry trade anomaly, when currencies with high interest rates and high inflation rates may appreciate relative to currencies with lower interest and inflation rates, currency prices tend to move slowly away from the long-term value, as predicted by purchasing power parity (PPP). When carry trades fail, currencies move very quickly toward the long-run PPP value.

There are three versions of PPP: the law of one price, absolute PPP, and relative PPP. The **law of one price** states that, absent transaction and transportation costs, the same item should have the same price in all countries, adjusted using current exchange rates. The most famous illustration of the law of one price is that of the Big Mac Index. According to the Big Mac Index, in 2008, the euro was overvalued by 47% against the U.S. dollar (USD) (i.e., a Big Mac sold in Europe at a price 47% greater than its price in the United States when converted into the same currency), while many Asian currencies were extremely cheap vis-à-vis the dollar. However, one could argue that these numbers are not necessarily reliable, because they are calculated on the basis of a single perishable good that cannot be bought and sold by all market participants at the same price.

Rather than using just one good to determine the value of currencies, **absolute PPP** compares the price of a basket of goods across countries and states that the basket of goods should have the same price in all countries, adjusted using current exchange rates. Absolute PPP is a much more reliable indicator of the exchange rate that would equalize the purchasing power of one currency (amount of goods and services that can be purchased with the currency) with that of another currency. One could then compare this PPP level with the current exchange rate level to identify possible future currency movements. For instance, the International Comparison Program of the United Nations collects data on the prices of goods and services for virtually all countries in the world and publishes PPP exchange rates for all currencies. Comparing these PPP levels with current spot levels allows one to calculate the deviation from PPP and thus rank currencies based on their level of expected overvaluation or undervaluation.

Another interesting currency relationship is defined by **relative PPP**, which provides a one-to-one link between inflation differential and exchange rate changes. Essentially, relative PPP states that over time, the change in the exchange rate between two countries should reflect the relative changes between local prices or, equivalently, the difference between the inflation rates of the two countries. Mathematically, this can be written as follows:

$$\frac{e_1}{e_0} = \frac{1 + \text{Inflation}_{\text{Domestic}}}{1 + \text{Inflation}_{\text{Foreign}}} \tag{28.15}$$

where e_t is the foreign exchange rate at time t (value of one unit of foreign currency in terms of the domestic currency).

APPLICATION 28.3.10

Assume that the domestic inflation rate of 5% per period exceeds the foreign inflation rate of 3%, and that the current exchange rate is that 2 foreign currency units equals 1 foreign currency unit. What is the end-of-period exchange rate consistent with relative purchasing power parity?

Answer: 2.039. The ratio of (1 + inflation rate) in Equation 28.15 is (1.05/1.03), which is 1.0194. The future exchange rate is found as 1.0194 times the current exchange rate of 2, which equals 2.039. Note that 2.039 units of domestic currency diminished by domestic inflation (divided by 1.05) is twice the value of 1 unit of foreign currency diminished by foreign inflation (divided by 1.03)—and so it is in parity with the current exchange rate of 2.

How can global macro managers use the relationship of relative PPP? Say, for instance, that the current spot exchange rate is USD 1.3600 = EUR 1, and that the anticipated inflation rates for the next year are 3% in the United States and 5% in Europe. Therefore, 1.3600 × 1.03/1.05 = USD 1.3341 = EUR 1. Thus, the market values imply that the euro will depreciate against the U.S. dollar to 1.3341 dollars per euro. A global manager can compare the exchange rate implied by relative PPP to the exchange rate that the manager forecasts. The next issue is to determine the timing and to identify a potential catalyst for this to happen.

Volatility models of currency trading typically involve the trading of options. Managers can profit from long option positions in times of rising volatility, and from short option positions during times of declining volatility. Managers who are short options can also profit from the time decay of options during times of stable volatility, as the prices of options decline as time passes and expiration approaches.

28.3.11 Risk Management and Portfolio Construction

In their youth, global macro funds were primarily one-person shops placing directional bets with a lot of leverage and very few risk controls. Their volatility was extremely high, and large losses were frequent. For example, the Quantum Fund gained $1 billion against the British pound in 1992 but lost $2 billion during the Russian crisis of 1998. However, this old-style school of global macro fund gradually disappeared after the 1990s. Today's global macro managers still enjoy a high degree of flexibility, but risk management and a disciplined investment approach have become essential components of their activities.

The risk management culture has definitely changed the way global macro strategies are implemented. Most modern global macro managers aim to optimally diversify their portfolio holdings in order to reduce and control risk. In doing so, they often use a combination of value at risk (VaR) measures and stop-loss orders. The former quantifies the estimated loss at different levels of probability and time horizons, and has the advantage of being applicable across all asset classes and instruments as well as at the portfolio level. It is used to allocate risk capital across trade ideas and traders. The latter, stop-loss orders, are intended to impose rational and disciplined behavior, forcing a manager to exit from losing trades regardless of conviction. As summarized by Bruce Kovner (the retired founder of Caxton Associates LP), stop losses should be set "at a point that, if reached, will reasonably indicate that the trade is wrong, not at a point determined primarily by the maximum dollar amount you are willing to lose.[7]" Stop-loss orders, though, are not guaranteed to be executed at the stop-loss price, so losses can far exceed expectations during turbulent market conditions.

28.4 HISTORICAL PERFORMANCE OF DIRECTIONAL STRATEGIES

This section contains a brief analysis of the historical performance of the CISDM Equity Long/Short Index and the CISDM Global Macro Index. Both indices cover the period from December 1989 to December 2014—that is, 301 monthly returns. Over this period, the Equity Long/Short Index had an annualized return of 10.7% with an annualized standard deviation of 7.8%; the Global Macro Index had an annualized return of 9% with a lower volatility of 5.5% (see Exhibit 28.3).[8] The correlation between the two monthly series was 0.56. Clearly, some diversification benefits would accrue to a portfolio that held both types of strategies.

For instance, an equally weighted portfolio of both indices results in a portfolio with annualized returns of 9.8% and an annualized standard deviation of 5.9%.

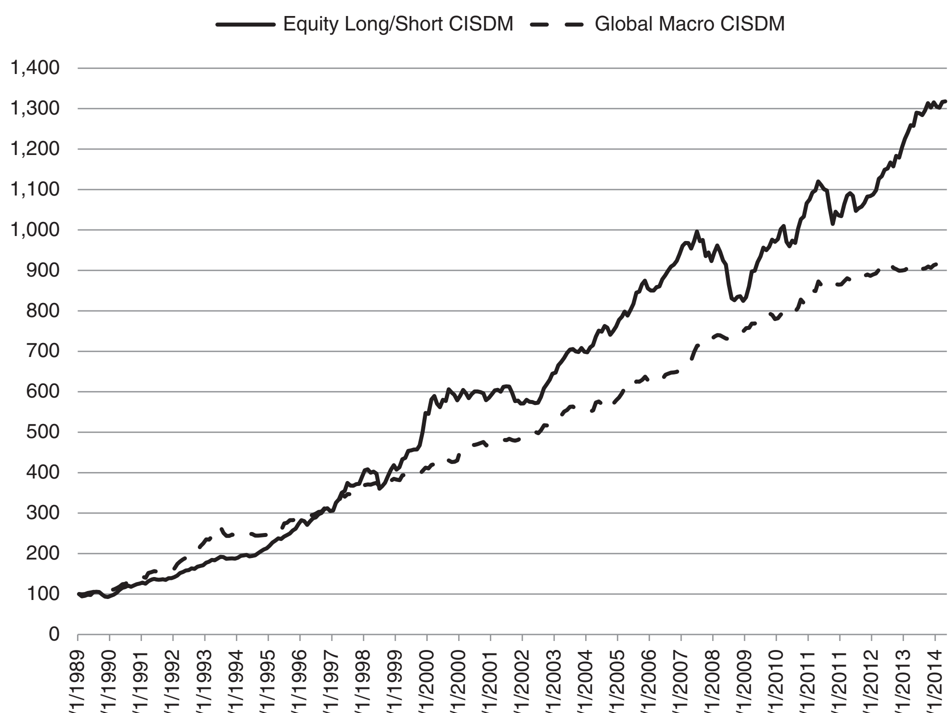


EXHIBIT 28.3 Historical Performance of Equity Long/Short and Global Macro Indices, 1989–2014

Such a portfolio would result in a higher Sharpe ratio (assuming a 3% annualized risk-free rate) of 1.15 [i.e., (9.8% – 3%)/5.9%], compared to a Sharpe ratio of 0.99 [i.e., (10.7% – 3%)/7.8%] for the Equity Long/Short Index and that of 1.09 [i.e., (9% – 3%)/5.5%] for the Global Macro Index.

28.5 CONCLUSION

Many argue that equity long/short and global macro are skill-based strategies. Clearly, both strategies take more concentrated positions when compared to equity market-neutral strategies. This chapter explored some of the more advanced topics that are well-known across the industry and possibly employed at some point by such managers. We also highlighted some of the investment procedures implemented by both managers. Equity long/short managers apply fundamental valuation principles in processing raw firm data. Moreover, they tend to take a very distinct approach toward understanding the companies that they invest in. Additionally, macro managers are well-versed in macroeconomic theories, such as PPP and interest rate parity conditions. While both strategies are unique in their underlying investment processes and procedures, their commonality stems from the managers taking concentrated discretionary bets while distilling distinct sets of informational variables.

NOTES

1. Eugene Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25, no. 2 (May 1970): 383–417.
2. Norris (2008).
3. For data on the performance of value versus growth stocks as well as return to a momentum strategy, see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.
4. For a very sound and useful resource on this approach, refer to the website run by Professor Aswath Damodaran of New York University: http://pages.stern.nyu.edu/~adamodar/.
5. At their peaks, Soros and Robertson had assets of about $22 billion. Soros Fund Management announced a revamping of the Quantum Fund amid steep losses in technology stocks, billions in redemptions, and the departure of two top managers: Stanley Druckenmiller and Nicholas Roditi. Robertson announced in March 2000 that he was liquidating his funds and closing the doors on Tiger Management after his bets on value stocks backfired.
6. Euroskeptics like to call it White Wednesday.
7. Quoted in J. D. Schwager, *Market Wizards* (Hoboken, NJ: John Wiley & Sons, 2013).
8. Annualized returns are computed by taking the average monthly returns and multiplying by 12. Annualized standard deviation is computed by taking the standard deviation of monthly returns and multiplying the results by the square root of 12.

REFERENCES

Ahl, P. 2001. "Global Macro Funds: What Lies Ahead?" *AIMA Newsletter*, April.

Ahmed, P. 2013. "Stock Market Efficiency and Market Microstructure in Emerging Markets." In *Market Microstructure in Emerging and Developed Markets: Price Discovery, Information Flows, and Transaction Costs*, edited by H. K. Baker and H. Kiymaz. Hoboken, NJ: John Wiley & Sons.

Alpert, M., and H. Raiffa. 1982. "A Progress Report on the Training of Probability Assessors." In *Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky, 294–305. Cambridge, MA: Cambridge University Press.

Asness, C., A. Frazzini, and L. Pedersen. 2014. "Quality Minus Junk." Working paper, AQR Capital Management.

Baker, M., and J. Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance* 61 (4): 1645–80.

———. 2007. "Investor Sentiment in the Stock Market." Working paper, National Bureau of Economic Research. www.nber.org/papers/w13189.

Barber, B., and T. Odean. 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investing." *Quarterly Journal of Economics* 116 (1): 261–92.

Barberis, N., and R. Thaler. 2003. "A Survey of Behavioral Finance." In *Handbook of Economics of Finance*, edited by G. Constantinides, M. Harris, and R. Stulz. Amsterdam: North-Holland.

Beer, A. 2015. "The Value Factor Reconsidered." Working paper, Beachhead Capital Management. www.beachheadcapital.com/Articles/The%20Value%20Factor%20Reconsidered.pdf.

Black, K. 2004. *Managing a Hedge Fund.* New York: McGraw-Hill.

Brav, A., W. Jiang, F. Partnoy, and R. Thomas. 2008. "Hedge Fund Activism, Corporate Governance and Firm Performance." *Journal of Finance* 63 (4): 1729–75.

Brown, S., W. Goetzmann, and J. Park. 2000. "Hedge Funds and the Asian Currency Crisis of 1997." *Journal of Portfolio Management* 26 (4): 95–101.

Burnside, C., M. Eichenbaum, and S. Rebelo. 2011. "Carry Trade and Momentum in Currency Markets." NBER Working Paper 16942.

Duong, C., G. Pescetto, and D. Santamaria. 2010. "Fundamental Analysis in Value-Glamour Contexts." Working paper, Canterbury Christ Church University, UK.

Eichengreen, B. J., A. Jansen, B. Chadha, L. Kodres, D. Mathieson, and S. Sharma. 1998. "Hedge Funds and Financial Market Dynamics." *IMF Occasional Papers* 166, International Monetary Fund.

Fama, E. 1975. "Short-Term Interest Rates as Predictors of Inflation." *American Economic Review* 65 (3): 269–82.

———. 1981. "Stock Returns, Real Activity, Inflation and Money." *American Economic Review* 71 (September): 545–65.

Fama, E., and K. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.

———. 1995. "Size and Book-to-Market Factors in Earnings and Returns." *Journal of Financial Economics* 50 (1): 131–56.

———. 2008. "Dissecting Anomalies." *Journal of Finance* 63 (4): 1653–78.

Financial Stability Board, 2000, "Report of the Working Group on Highly Leveraged Institutions," Bank of International Settlements, Switzerland

Fischhoff, B., P. Slovic, and S. Lichtenstein. 1977. "Knowing with Certainty: The Appropriateness of Extreme Confidence." *Journal of Experimental Psychology: Human Perception and Performance* 3 (4): 552–64.

Fisher, I. 1930. *The Theory of Interest, as Determined by Impatience to Spend Income and Opportunity to Invest It*. New York: Macmillan.

Frazzini, A., and L. Pedersen. 2013. "Betting against Beta." *Journal of Financial Economics* 111 (1): 1–25.

Fung, W., D. Hsieh, and K. Tsatsaronis. 2000. "Do Hedge Funds Disrupt Emerging Markets?" In *Brookings-Wharton Papers on Financial Services*, edited by R. Litan and A. Sanatomoera, 377–421. Washington, DC: Brookings Institution.

Harvey, C., Y. Liu, and H. Zhu. 2016. "… and the Cross-Section of Expected Returns." *Review of Financial Studies* 29 (1): 5–68.

Kahneman, D., and A. Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–91.

Liew, J., and G. Wang. 2015. "Twitter Sentiment and IPO Performance: A Cross-Sectional Examination." Working paper, Johns Hopkins Carey Business School.

Menkhoff, L., L. Sarno, and A. Schrimpf. 2012. "Currency Momentum Strategies." *Journal of Financial Economics* 106 (3): 660–84.

Norris, F. 2008. "Porsche Reinvents the Short Squeeze." *New York Times*, October 30. www.nytimes.com/2008/10/30/business/worldbusiness/30iht-norris31.1.17372644.html.

Pedersen, L. H. 2015. "Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined." Princeton, NJ: Princeton University Press.

Pojarliev, M., and R. Levich. 2008. "Do Professional Currency Managers Beat the Benchmark?" *Financial Analysts Journal* 64 (5): 18–32.

Safvenblad, P. 2003. "Global Macro and Managed Futures Strategies." Working paper, Stockholm School of Economics.

Seddighi, H. R., and W. Nian. 2004. "The Chinese Stock Exchange Market: Operations and Efficiency." *Applied Financial Economics* 14 (11): 785–97.

CHAPTER 29

Hedge Funds: Credit Strategies

This chapter discusses hedge fund strategies involving market-traded fixed-income instruments that are subject to credit risk. The issuers of these instruments can be either corporate or sovereign borrowers, and the instruments typically bear lower than investment-grade ratings. The lower rating implies that there is a significant default risk associated with these instruments, which partially explains the higher yield and the premium that investors can expect for investing in these instruments. Hedge funds investing in these types of instruments employ such strategies as distressed debt investing and asset-backed lending.

The first part of the chapter discusses the nature of credit risk and presents three approaches to credit risk modeling. While these models can be very complex, our goal is to provide readers with a basic understanding of these models and highlight the strengths and weaknesses of each approach.

The second part of the chapter discusses two hedge fund strategies that attempt to exploit inefficiencies of markets for certain types of credit instruments that can add value to an investor's portfolio. The first strategy discussed is distressed debt investing, which is well known and has a relatively long history. The second strategy is asset-based lending, which is a relatively new strategy in the hedge fund space.

29.1 THE ECONOMICS OF CREDIT RISK

Credit risk is the risk of loss resulting from some type of credit event with a counterparty. **Credit** is money or funds granted by a creditor or lender to a debtor or borrower. The types of credit issued could be bank loans, corporate bonds, or government (sovereign) bonds. For a variety of reasons, the debtor may not be able or willing to meet all the obligations that the credit contract stipulates. For example, if the debtor does not have the funds to make the regular interest payments on a loan, the debtor will be in default. However, credit risk could arise from a set of credit events, of which nonpayment of obligations is just one example. **Credit events** that give rise to credit risk include bankruptcy, downgrading, failure to make timely payments, certain corporate events, and government actions:

- BANKRUPTCY: This arises if an entity is dissolved or becomes insolvent and therefore is unable to meet its obligations.
- DOWNGRADING OF CREDIT RATING: This arises when the external credit-rating agencies lower the credit rating of an entity due to changes in the financial condition of the entity or changes in overall economic conditions.

- Failure to make timely payments: A borrower may fail to make timely interest or principal payments even if the entity is not dissolved and does not face economic hardships.
- Corporate events: Certain corporate events, such as mergers or spin-offs, could weaken the financial condition of a firm, making it more difficult for the entity to meet its financial obligations.
- Government actions: Capital controls and other government restrictions could prevent a borrower from meeting its financial obligations.

The extent and the consequence of credit risk depend on a number of factors, of which credit exposure or exposure at default (EAD) is the most important. EAD measures the potential loss to the creditor in the case of a credit event. Closely related to EAD is loss given default (LGD), which takes into account any potential recovery should default take place. In other words, if 100% of EAD is lost in a case of default, then LGD = EAD. However, in most cases the creditor is likely to recover some of the losses; therefore, LGD is typically less than EAD.

29.1.1 Adverse Selection and Credit Risk

Adverse selection refers to an economic process in which negative or undesirable outcomes take place when the parties to a transaction have asymmetric information (i.e., they have access to different information).[1] It is often the case that borrowers have more information than lenders about the borrowers' ability and willingness to meet their obligations. In this case, there is asymmetric information between borrowers and lenders, with borrowers having more information about their circumstances than lenders. This asymmetric information leads to adverse selection.

Since the lender is at a disadvantage, it may decide to raise the rate of interest it would charge potential borrowers. However, this will have some unintended consequences. As the cost of borrowing increases, a disproportionate fraction of borrowers willing to pay the high cost of borrowing will be privately aware of their own poor credit quality and will therefore find borrowing at such a high cost still attractive. If the lender were to raise the cost of borrowing any further in order to compensate for this adverse selection, the proportion of poor credit quality would continue to increase. At the extreme, the very high interest rate will drive all borrowers away, and the lender will be left with unused capital.

Financial institutions have developed a variety of tools and processes to reduce the impact of adverse selection. In addition to taking into account a borrower's credit history and reputation, lenders may ask for collateral, since it is often easier to verify the economic value of collateral. Also, lenders may limit the size of their loans to an individual or a group of borrowers in order to diversify their risk and reduce the impact of adverse selection.

29.1.2 Moral Hazard and Credit Risk

Whereas adverse selection is a potential problem before a financial transaction is completed, moral hazard can come into play after the transaction is completed, such as when one party to an economic transaction takes actions or takes on more risks while the counterparty bears the consequences. For example, after a loan transaction has been completed, the corporate borrower may decide to use the loan's proceeds to pay large dividends to its shareholders. This action may weaken the financial condition of the borrower and therefore increase the risk of default faced by the lender. Another important example is the case of financial lending institutions that, because their deposits are insured by a government agency, can raise money at a very low rate of interest. Moral hazard will arise if these financial institutions are allowed to use the funds they have raised at low interest rates to make risky loans and earn a large spread. If the borrowers default on their loans, governments and their citizenry will bear the cost of those defaults. To reduce moral hazard, lenders monitor the behavior of borrowers, impose restrictions on how the loans' proceeds may be used, and limit the size of their loans to risky borrowers.

29.1.3 Probability of Default

As previously noted, EAD and LGD are two important factors determining the extent and the consequence of credit risk. Another important factor that affects credit risk is the probability of default (PD), which refers to the probability that the borrower will be unable to meet its financial obligations. The task of assigning a default probability to each borrower in a lender's credit portfolio is a difficult task, as it is affected by a number of firm-specific as well as macroeconomic conditions. Adverse selection and moral hazard impact the probability of default; thus, any actions taken to mitigate their impacts are done with the aim of reducing the probability of default.

Lenders use their own experience as well as market data and credit ratings to improve the accuracy of their estimates of PD. As discussed in the next section, certain credit risk models use market data such as credit spreads to improve estimates of PD. Credit spreads refer to the risk premium that instruments exposed to credit risk must offer lenders in order to meet the credit needs of borrowers. A credit model that relates credit spreads to PD can be used to deduce the PD implied by the observed credit spreads. Alternatively, lenders can use credit ratings assigned by external credit rating agencies to improve their estimates of PD. Using historical data available from these agencies, one can obtain empirical estimates of relative numbers of defaults for each credit rating. For example, we may observe that 0.15% of firms rated Baa1 have historically defaulted on their loans. Assuming that these historical relationships are relatively stable, one can use current rating of a Baa1 bond to conclude that its PD is 0.15%.

29.1.4 Expected Credit Loss

EAD and LGD are related to each other in terms of the recovery rate (RR), which gives the percentage of EAD that can be recovered through legal and economic processes. Typically, the recovery does not take place at the time of default; indeed, it may take years to gain control of the recovered amount. In such cases, the present value of the recovered amount must be used to calculate the recovery rate. Note that the recovery rate can be expressed as:

$$\text{Present Value of Sum to Be Recovered/EAD}$$

The EAD can be expressed as the sum of any principal and interest due.

Example: Suppose the EAD of a loan due at the end of this year is \$100 million. Should there be a default, the lender expects to recover 40% of the principal after five years. If the appropriate annual rate of interest for discounting the cash flows of the same riskiness as the recovered amount is 6%, the actual recovery rate is:

$$\text{Sum to be recovered} = 0.4 \times 100 = 40 \text{ million}$$

$$\text{Present value of sum to be recovered} = \frac{40 \text{ million}}{(1 + 0.06)^5} = 29.9 \text{ million}$$

$$\text{Recovery rate} = \frac{29.9}{100} = 29.9\%$$

Given the definitions of EAD, LGD, RR, and PD, we can obtain an estimate of loss and expected loss from credit risk. In particular,

$$\text{LGD} = \text{EAD} \times (1 - \text{RR}) \tag{29.1}$$

$$\text{E}[\text{Loss}] = \text{LGD} \times \text{PD} = \text{EAD} \times (1 - \text{RR}) \times \text{PD} \tag{29.2}$$

APPLICATION 29.1.4

Suppose a loan has \$100 million of principal and \$10 million of interest due at the end of the year. If there is a default, the lender expects to recover \$25 million three years after the default. The PD for the borrower is estimated to be 1%. Find the exposure at default, the present value (PV) of the sum to be recovered, the recovery rate, the loss given default, and the expected loss using an interest rate of 8%.

These values are given as follows:

$$\text{EAD} = \$100 \text{ million principal} + \$10 \text{ million interest} = \$110 \text{ million}$$

$$\text{PV of sum to be recovered} = \$25 \text{ million}/1.08^3 = \$19.8458 \text{ million}$$

$$\text{RR} = \text{PV of sum to be recovered}/\text{EAD} = \$19.8458 \text{ million}/\$110 \text{ million} = 0.1804164$$

$$\text{LGD} = \text{EAD}(1 - \text{RR}) = \$110 \text{ million } (1.0 - 0.1804164) = \$90.1542 \text{ million}$$

$$\text{Expected loss} = \text{LGD} \times \text{PD} = \$90.1542 \text{ million} \times 1\% = \$0.901542 \text{ million}$$

Note that LGD can also be found as EAD – Present value of sum to be recovered.

29.2 OVERVIEW OF CREDIT RISK MODELING

In constructing a fixed-income portfolio, an investor faces the decision of how much, if any, interest rate risk or credit risk should be taken. Owners of government bonds, such as U.S. Treasuries or German bunds, predominantly face interest rate risk, as the risk that the U.S. or German government would default on its obligations is considered to be extremely low.

However, owners of bonds issued by levered entities (i.e., borrowers with significant amounts of leverage on their balance sheets) face a different risk profile altogether, as the main risk in those instruments lies in the risk that the credit quality of the borrower will deteriorate, resulting in a credit event taking place. As discussed earlier, default is one extreme example of a credit event, one in which the firm is unable to continue operating as a going concern. In practice, default is not a terminal end—that is, a company does not cease to exist at the point of default. In fact, the process of workout or restructuring is one in which the underlying business practices, management, and liabilities of the borrower are significantly changed in order to enable the business to emerge as a going concern or otherwise face liquidation. Nonetheless, the focus on many credit risk models assumes default as an end point.

To invest in instruments that are exposed to credit risk and to take advantage of investment opportunities offered by such instruments, investors must have a thorough understanding of the credit risk and credit exposures of these instruments. For instance, in a relative value strategy, an investor may decide to compare returns on two instruments with different levels of exposure to credit risk to determine if both offer the appropriate credit spreads given their levels of credit risk. A credit risk model can be used to evaluate both instruments in order to determine whether the instruments are correctly priced relative to each other.

There are three types of credit risk modeling approaches: the structural approach, the reduced-form approach, and the empirical approach.

In the structural approach, the framework is set around an explicit relationship between capital structure and default. The value of a firm's assets is set equal to the value of its equity plus the value of its debt. Equity of the firm is viewed as a call option on the firm's assets, with the strike price being the face value of the debt due at the time of exercise. In contrast, bondholders are viewed as having a risk-free bond and a short position in a put option on the firm's assets. If the value of assets is less than the face value of the debt, the put option will be exercised on the bondholders, resulting in their giving up the risk-free bond and receiving the firm's assets.

In the reduced-form approach, default is modeled as an exogenous event that is driven by a random signal. The behavior of this random signal is the deciding factor of default rather than the value or the dynamics of the firm's assets. Reduced-form models are built on the assumption that default is a random event that can be described using statistical and economic models.

The **empirical approach to credit risk modeling** is based on the assumption that it is too difficult to model the company and its environment accurately. Instead, an investor looks at companies that have defaulted and evaluates their financial data to understand the credit risk of the firm. This approach produces a credit score that is used to rank firms in terms of their credit-worthiness.

29.3 THE MERTON MODEL

The best-known of the structural modeling approaches is Merton's structuring model (see Merton 1974), which was discussed in Chapter 25 of CAIA Level I.

29.3.1 Capital Structure in the Merton Model

The approach starts with a simple capital structure of Assets = Liabilities (Debt) + Equity.

$$A_t = D_t + E_t \tag{29.3}$$

It assumes that default happens at the maturity of the debt if the asset value falls below the face value of the debt. Other assumptions of this model are that there is no cost to bankruptcy, that debt and equity can be traded without friction, and that the debt is a zero-coupon bond with the face value of K and a maturity date of T. Equity value at time T is:

$$E_T = \max(A_T - K, 0) \tag{29.4}$$

Note that this payoff is the payoff of a European call option on underlying assets A_T, strike K, and maturity T. The payoff to bondholders at time T is:

$$D_T = K - \max(K - A_T, 0) \tag{29.5}$$

This means that bondholders will receive the face value unless the face value is less than the value of the assets. In that case, bondholders will receive the assets. In other words, the equity holders first pay creditors the face value of the debt. Then they exercise the put option given to them by creditors, receiving back the face value in exchange for the firm's assets.

29.3.2 The Merton Model and the Black-Scholes Option Pricing Model

Since the Merton model views capital structure with options, the model can be applied using option pricing models, such as the Black-Scholes option pricing model (Black and Scholes 1973), to price a firm's equity as a call option on the assets of the underlying company. This use of a well-known option model is intuitive and one of the most appealing aspects of the Merton model. Using the Black-Scholes formula for pricing a European call option results in:

$$E_t = A_t \times N(d) - K \times e^{-r\times\tau} \times N\left(d - \sigma_A\sqrt{\tau}\right) \tag{29.6}$$

Here, r is the annualized short-term rate of interest on risk-free debt, $\tau = T - t$ is the time left to maturity of the debt, σ_A is the annualized volatility of the rate of return on the firm's assets, and

$$d = \frac{\ln(A_t/K) + \left(r + 0.5\sigma_A^2\right) \times \tau}{\sigma_A\sqrt{\tau}} \tag{29.7}$$

The symbol $N(\cdot)$ refers to the cumulative probability distribution function for a standard normal distribution. That is, $N\left(d\right)$ = Probability $\left(Z \leq d\right)$, where Z is a standard normal random variable.

In this scenario, debt holders face a potential loss of the shortfall between the asset value A_T and the face value of the debt K (note that the equity holders cannot lose more than their equity stake). Thus, if debt holders were to buy a put option with a strike K that pays off if asset value A_T falls below K, the debt holders would be hedged against the loss in the event of default. Thus, at any given time t, the debt and the put option guarantee a payoff of K at maturity time T.

As previously discussed, the price of a risky debt issue at any time is equal to the price of a portfolio that consists of a long position in risk-free debt with the same maturity as the original debt, and a short position in a put option with a strike price equal to the face value of the bond and the same maturity date as the bond. That is,

$$D_t = K \times e^{-r\times\tau} - (\text{Put Price})_t \tag{29.8}$$

Using the Black-Scholes formula for a European put option, the put price is given by:

$$P_t = K \times e^{-r\times\tau} \times N\left(-d + \sigma_A\sqrt{\tau}\right) - A_t \times N(-d) \tag{29.9}$$

29.3.3 The Role of the Credit Spread in the Structural Model

At this point, we are ready to present a model of the pricing of the risky debt in terms of the credit spread. The common practice in valuing risky debt is to account for the credit risk in terms of a spread above the risk-free rate. In the current context, the value of the zero-coupon debt can be expressed as:

$$D_t = K \times e^{-(r+s_t)\times\tau} \tag{29.10}$$

where s_t is the annual spread due to credit risk. Combining Equations 29.8, 29.9, and 29.10, we can obtain an expression for the credit spread as it relates to the parameters of the Merton model.[2]

$$s_t = -\frac{1}{\tau} \times \ln\left[N\left(d - \sigma_A\sqrt{\tau}\right) + \frac{A_t}{K}e^{r\times\tau} \times N(-d)\right] \tag{29.11}$$

Example: The ABC Corporation has the following balance sheet: Its assets are worth €100 million, and it has four-year maturity zero-coupon debt with a face value of €70 million. The volatility of its assets is 20% per year, and the risk-free rate is 5% per year. We want to calculate the current values of ABC's debt and equity. We need to first use Equation 29.7 to calculate d:

$$d = \frac{\ln(100/70) + \left(0.05 + 0.5 \times 0.2^2\right) \times 4}{0.2 \times \sqrt{4}} = 1.592$$

$$d - 0.2 \times \sqrt{4} = 1.192$$

Then, we calculate the area under the standard normal distribution for the value of 1.592 and 1.192. That is, $N(1.592) = 0.944$ and $N(1.192) = 0.883$. We are now prepared to apply Equation 29.6:

$$E_t = 100 \times 0.944 - 70 \times e^{-0.05\times4} \times 0.883 = 43.79$$

This means that the current value of the debt must be:

$$D_t = 100 - 43.79 = 56.21$$

We can also calculate the value of the put option given to equity holders. As previously stated, the value of the put option given to equity holders is equal to the difference between current value of the risk-free debt and the risky debt:

$$\begin{aligned}\text{Put Option} &= \text{Risk-Free Debt} - \text{Risky Debt}\\ &= (70 \times e^{-0.05\times4}) - 56.21 = 1.1\end{aligned}$$

Therefore, the value of the put option given to equity holders is €1.1 million. Finally, we can see that the credit spread is:

$$s = -\frac{1}{4} \times \ln\left[0.883 + \frac{100}{70} \times e^{-0.05\times4} \times (1 - 0.944)\right] = 0.49\%$$

We can verify that the spread correctly prices the risky debt. That is, 56.21 is the present value of 70 discounted back at 5.49% for four years.

29.3.4 Evaluation of the Merton Model

In its simplest form, the Merton model is just the starting point of credit risk modeling. Because its underlying assumptions are highly restrictive, it cannot be applied to real-world situations in which firms have different types of coupon bonds, and interest rates are random through time. Still, the Merton model has many intuitive properties and is a useful point of departure for more complex models. For example, it can be shown that the risk-neutral probability of default can be expressed as $\Pr[A_T \leq K] = 1 - N(d)$. It is important to note that this is not the true probability of default; rather, it is a version of that probability implied by current prices if those prices were determined in markets where investors were risk neutral. It is equally important to realize that even if investors were risk neutral, they would still care about credit risk and loss of principal. However, risk-neutral investors do not demand an extra risk premium for systematic risk (i.e., the beta of the asset in the capital asset pricing model [CAPM]) of the investment that contains credit risk.

The shortcomings of the Merton model were discussed in Chapter 25 of CAIA Level I, and it was pointed out that some of the parameters of the model are not readily observable—particularly the market value of assets and the return volatility. The KMV model, discussed in the next main section, attempts to overcome this difficulty. From an empirical point of view, the Merton model performs poorly in explaining the credit spread on short-term securities. In other words, if the parameters of the model are selected so that they correspond to available historical data, the result would predict a very low credit spread for short-term assets, results that are contradicted by available empirical evidence.

29.3.5 Four Important Properties of the Merton Model

In this section, we take a closer look at some of the important properties of the Merton model. As discussed so far, two important outputs from the Merton model are the credit spread and the probability of default. Here we examine how these two outputs are affected if there are changes in some of the inputs to the model. For example, what will happen to the credit spread if the maturity of the debt increases? As we will see shortly, the answer to this question is neither obvious nor simple.

To facilitate our discussion, let's recall the two expressions that the Merton model provides for the two important outputs:

$$\text{Credit spread: } s_t = -\frac{1}{\tau} \times \ln\left[N\left(d - \sigma_A\sqrt{\tau}\right) + \frac{A_t}{K}e^{r\times\tau} \times N(-d)\right] \quad (29.12)$$

$$\text{Probability of default: } \Pr[A_T \le K] = 1 - N\left(d\right) \quad (29.13)$$

$$d = \frac{\ln(A_t/K) + \left(r + 0.5\sigma_A^2\right) \times \tau}{\sigma_A\sqrt{\tau}} \quad (29.14)$$

The following are four important properties of the Merton model:

1. Sensitivity to maturity: The probability of default increases as the time to maturity increases. This appears to be an intuitive result, because as the time to maturity increases, the firm's assets have more chances to decline below the face value of the debt. It must be noted that the probability of default increases at a decreasing rate when maturity is extended. For example, if the probability of default increases by 1% as the maturity is extended from two to three years, the probability will increase by less than 1% when maturity is extended from six to seven years.

The change in credit spread as maturity increases is less obvious. It can be shown mathematically that the credit spread increases initially but then begins to decline slightly as maturity increases. The fact that the credit spread should increase initially makes sense. We just learned that the probability of default increases as the time to maturity increases. However, what can explain the decline in the credit spread as maturity increases further? To understand the reason behind this result, let's remember that risky debt is really a portfolio consisting of risk-free debt and a short position in a put option on the firm's assets. Because this put can be exercised only on the maturity date of the debt, it is a European option. While the value of an American put option will always increase as its maturity increases, the same cannot be said of a European put option. For example, if it is optimal to exercise a put option today, the value of the put option declines if the owner of the put option is forced to delay its exercise. The same phenomenon affects the value of the put option owned by shareholders. As the maturity date of the put option exceeds the optimal exercise date, the value of the put option begins to decline, reducing the credit spread that must be used to calculate the price of the risk debt (note that the risky debt is short the put option).

2. Sensitivity to asset volatility: As expected, the probability of default increases as the volatility of the asset increases. Again, similar to an increase in maturity, the default probability increases at a decreasing rate. The credit spread will also increase as the volatility of the asset increases. However, the credit spread may increase by a smaller amount for longer-maturity bonds compared to short-term bonds. The reason for this is that as volatility increases, there is a greater chance for the value of assets to end up very far above or below the face value of the debt. However, there is a lower bound to the value of assets that cannot become negative. For this reason, at very high levels of asset volatility, the increase in credit spread will be smaller for long-term bonds. That is, the negative impact of the higher volatility is limited to the downside, while the positive impact of higher volatility is unlimited to the upside.

3. Sensitivity to leverage: As leverage increases, both the probability of default and the credit spread increase. While the probability of default increases by a larger amount for long-term bonds as leverage increases, the credit spread may increase by a relatively smaller amount for long-term bonds. The reason behind this reaction is the same as the one discussed in the context of the sensitivity to increases in asset volatility.

4. Sensitivity to the riskless rate: How default probability and credit spreads react to changes in the short-term riskless rate depends very much on how the rate of return on the asset is assumed to react as the riskless rate changes. The Merton model implicitly assumes that the expected rate of return on the firm's assets is equal to the riskless rate plus a constant risk premium. Under this assumption, as the riskless rate increases, the mean return on the firm's assets increases, reducing the probability of default and the credit spread. However, if this assumption is relaxed, then the reaction to an increase in the riskless rate will be difficult to predict.

29.4 OTHER STRUCTURAL MODELS—KMV

Another variation of a structural model that is widely implemented in the industry is the KMV model. KMV stands for Kealhover, McQuown, and Vasicek, who founded the firm KMV in 2002 (Crosbie and Bohn 2003). The firm was sold to Moody's Investors Service, and the model is now the basis of Moody's modeling approach.

29.4.1 Overview of the KMV Credit Risk Model

The **KMV model** is a structural credit risk model that uses Merton's model and estimates of the volatility and total value of the firm's underlying assets and equity to estimate the credit risk of the debt. The novelty of the KMV model is that it turns the creditor's lending problem around and considers the loan repayment incentive problem from the viewpoint of the borrowing firm's equity holders. As previously discussed, the value of assets, A_t, and its volatility, σ_A, cannot be observed or estimated directly. The KMV model attempts to solve this problem by estimating them simultaneously using the Merton model and the economic relationship between equity values and firm values. To solve for the two unknowns, the model uses (1) the structural relationship between the market value of a firm's equity and the market value of its assets (see Equation 29.6), and (2) the relationship between the volatility of a firm's assets and the volatility of a firm's equity. This relationship is expressed as:

$$\sigma_E = \frac{A_t}{E_t} \times \Delta \times \sigma_A \tag{29.15}$$

The term Δ is the delta of the equity with respect to changes in the value of the assets and, under the Merton model, is equal to $N(d)$. Notice that the market value of equity is observable for publicly traded firms, and the value of the equity return volatility can be estimated from historical equity returns. Once the values of E_t and σ_E are obtained, the two relationships previously stated can be used to solve for the two unknowns, A_t and σ_A. Two important outputs of the KMV model are the probability of default—or, using KMV's terminology, the expected default frequency—and a credit score known as distance to default. The next section will briefly discuss how these two figures are calculated.

An important difference between the KMV model and the Merton model is the way default is modeled. The KMV model uses a default trigger to model default. In the Merton model, the **default trigger** for the firm's total asset value is the face value of the zero-coupon bond, because it represents the asset value at which the firm would be on the brink of default. If the total value of the firm's assets falls below the face value of the debt, the firm is modeled as being in default. In practice, firms have short-term and long-term debt, and KMV takes a weighted average of face values of short-term and long-term debt to calculate the default trigger. KMV argues that short-term debt is more pressing, since the firm will have to service this debt, but it has some leeway on how to service longer-term debt. Thus, the default trigger is calculated when the value of the firm's assets falls below the full amount of the short-term debt and a partial amount of longer-term debt.[3]

29.4.2 Using the KMV Model to Estimate a Credit Score

As mentioned, one of the main outputs of the KMV model is a credit score, which is measured as distance to default for a borrower. The **distance to default (DD)** is approximately measured as the percentage difference between a firm's assets and its default trigger relative to the volatility of its assets:

$$\mathrm{DD}_t = \frac{A_t - K}{A_t \times \sigma_A} \tag{29.16}$$

Here, K is the default trigger of the firm, which would be the face value of the debt in the Merton model and a weighted average of the face values of short-term and long-term debt in the KMV model. For example, suppose that a firm's assets are valued at \$100 million and the default trigger is \$80 million, reducing the equation to DD = $0.2/\sigma_A$. The firm's assets must decline 20% in this example to trigger default. The annual standard deviation of the returns of the assets is σ_A. The resulting ratio (DD) is the number of standard deviations that the assets must lose in order to decline in value to reach the default trigger. Suppose that the standard deviation (σ_A) in this example is 0.1, so that DD = 2. The value of 2 indicates that this firm will enter the default region if the value of its assets drops by two standard deviations of its value—that is, by \$20 million. Thus, if the DD of a firm is reported to be n, then the firm will enter the default region if the value of its assets drops by $n \times \sigma_A$. Alternatively, the firm can be viewed as being n standard deviations away from default due to asset value declines.

APPLICATION 29.4.2

Suppose the assets of KYZ Corporation are worth \$200 million. Its default trigger is estimated to be \$120 million, and the volatility of its assets is 40% per year. What is KYZ's distance to default?

The DD of KYZ is found using Equation 29.16 as:

$$DD = \frac{200 - 120}{200 \times 0.4} = 1$$

KYZ Corporation is one standard deviation of asset volatility away from default.

29.4.3 Using the KMV Model to Estimate an Expected Default Frequency

As mentioned in the last section, an important output from the KMV model is the calculation of the **expected default frequency** (**EDF**), which will measure theoretically or empirically the probability that loans of certain characteristics could default. The value of DD can be used to ask the following question: What is the probability that the value of a firm's assets could drop by n standard deviations over the t periods? Here we discuss how the empirical EDF can be calculated.

Suppose that we have a large historical database of firm defaults and loan repayments, and we calculate that the firm we are analyzing has a distance to default of n (DD $= n$). We then ask the empirical question: What percentage of firms in the database actually defaulted within the one-year time horizon when their asset values placed them a distance of n standard deviations away from default at the beginning of the year, and how does that compare to the total population of firms that were n standard deviations away from default at the beginning of the year? This produces an empirical EDF:

$$EDF = \frac{\text{Number of Firms That Defaulted within One Year with DD} = n}{\text{Total Number of Firms in the Population with DD} = n} \quad (29.17)$$

APPLICATION 29.4.3

Suppose there were 500 firms with DD = 2 at the beginning of 2014 in a sample of European corporations. Of this total, 8 defaulted by the end of the year. Estimate the EDF empirically.

The estimate of the EDF is found using Equation 29.17:

$$\text{EDF} = \frac{8}{500} = 1.6\%$$

Based on historical observation, we might expect 1.6% of European firms with DD = 2 to default within one year. The percentage can be used to obtain other valuable estimates, such as the credit spread that bonds issued by such firms should command.

29.5 REDUCED-FORM MODELS

In a reduced-form model, the framework does not think about what causes default; that is, default is exogenous. Key drivers in a reduced-form model include time to default (or default time) and recovery in the event of default (conversely, loss given default). The difference between various reduced-form models typically involves the processes of modeling when default occurs and estimating the recovery if default happens.

29.5.1 Default Intensity in Reduced-Form Models

Under certain assumptions about the nature of the random process that leads to default, the probability of survival for a given number of years can be shown to have an exponential distribution. Assuming an exponential distribution and denoting $p(t)$ as the probability that a firm has survived for t years, this probability can be expressed as:

$$p(t) = \exp(-\lambda \times t) \tag{29.18}$$

where λ is called the **default intensity** of the model. The parameter λ determines both the expected time to default and the probability of survival. Specifically, given a default intensity of λ, the expected time to default is $(1/\lambda)$. This means the higher the default intensity, the shorter the expected time to default. The probability of default at or before t is given by $1 - p\,(t)$. The default intensity that appears in Equation 29.18 may refer to the actual or physical default intensity or to the risk-neutral version of the same variable. If the actual default intensity is used, then the analyst obtains an estimate of the actual expected time of default. The risk-neutral version of the default intensity is used exclusively for valuation purposes and to estimate credit spreads on investments exposed to credit risk.

Suppose default time is continuous, meaning that default can take place at any time, and not just at discrete points of time (e.g., at the end of each quarter). Then, given Equation 29.18, the probability of default in the time interval of $(t, t + \Delta t)$, assuming there has been no default up to time t, is given by $\lambda \times \Delta t$, where Δt is a relatively small length of time. This is the conditional probability of default. On the other hand, the unconditional probability of default in the time interval of $(t, t + \Delta t)$ is given by $\exp(-\lambda \times t)\,\lambda \times \Delta t$, which is equal to the probability of surviving up to time t multiplied by the conditional probability of defaulting between t and Δt. It is important to note that λ is not the probability of default within one year, as one year is a relatively long length of time. Finally, the probability that default could take place between s and t, assuming that no default has taken place up to time s, is given by:

$$p(s) - p(t) = \exp(-\lambda \times s) - \exp(-\lambda \times t) \tag{29.19}$$

For example, suppose a start-up company has an actual default intensity of 5%. At the time the firm is established (i.e., time 0), the probability that the firm will default within year 3, assuming that it has already survived for two years, is:

$$\begin{aligned}\text{Probability of Default in Year 3} &= \exp(-0.05 \times 2) - \exp(-0.05 \times 3)\\ &= 0.905 - 0.861 = 4.4\%\end{aligned}$$

That is, at time zero, there is a 4.4% chance that the firm will not survive beyond the third year, provided that it has already survived for two years. Also, the expected time to default is (1/0.05) = 20 years.

APPLICATION 29.5.1

Company A has an actual default intensity of 0.10, and Company B has a 17% probability of defaulting in the next three periods. Find (1) the probability of default, (2) the probability of survival, and (3) the expected time to default for Company A over the next two periods and the default intensity (λ) for Company B.

The expected time to default of any company is simply ($1/\lambda$), which in the case of Company A is 10 periods. Both remaining problems for company A are solved using Equation 29.18. For Company A, the probability of survival is $e^{-0.10\times2}$, which equals 0.819. The probability of default is simply 1.0 minus the probability of survival: 1.0 – 0.819, or 0.181. Company B has a probability of survival of 83%, or 0.83 (since its default probability is 17%). Therefore, the default intensity can be found by inserting 0.83 into the left side of Equation 29.18 and taking the natural logarithm of both sides. This generates the value that is equal to $-\lambda t$, specifically, ln(0.83), or –0.18633. Dividing that quantity by –3 (i.e., $-t$) generates a default intensity (λ) of 6.21%.

One of the contributions of a reduced-form model is its attempt to relate the default intensity to various financial and economic variables. Credit analysts have built models in which the default intensity is related to financial conditions of a firm as well as to macroeconomic conditions. Such a model is then typically calibrated by selecting its parameters so that it can explain historical default patterns as well as current conditions in credit markets (e.g., credit spreads for various credit ratings). The resulting model is then used to price new issues or to determine if certain credit instruments are priced too richly or too cheaply, which is crucial to the development of some credit investment strategies.

29.5.2 Valuing Risky Debt with Default Intensity

The default intensity model can be incorporated into the valuation of risky debt. To see the intuition of this model, let's consider a zero-coupon bond with the face value of K and time to maturity of T. If the bond is risk-free, then its current price is given by $K \times e^{-r\times T}$. Suppose the bond is exposed to default risk with a default intensity of λ. The probability of survival up to time T is $e^{-\lambda\times T}$, which means the probability of default by time T is $(1 - e^{-\lambda\times T})$. Suppose that in the case of default the bond will have zero recovery. Then the price is given by:

$$\begin{aligned} D_0 &= e^{-r\times T}(\text{Prob}_{\text{No Default}} \times K + \text{Prob}_{\text{Default}} \times 0) \\ &= e^{-r\times T}(e^{-\lambda\times T} \times K + (1 - e^{-\lambda\times T}) \times 0) = e^{-(r+\lambda)\times T}K \end{aligned} \tag{29.20}$$

APPLICATION 29.5.2

Suppose the riskless rate is 2%. What is the current price of a one-year zero-coupon bond issued by a start-up firm with a default intensity of 5%? The face value of the bond is $50 million.

The value ($46.62 million) can be found using Equation 29.20:

$$D_0 = e^{-(0.02+0.05)\times 1} \times 50 = \$46.62 \text{ million}$$

It is important to note that Equation 29.20 was obtained under the assumption that investors do not care about potential systematic risk of the bond, and the only risk that enters into their calculations is the default risk. In other words, the default intensity and the resulting probability of default are assumed to be under risk neutrality.

The calculation in Application 29.5.2 can be reversed, and using observed market prices, one can calculate the implied risk-neutral default intensity. For example, suppose the market price of the bond in the previous example is actually $45.24 million. The implied risk-neutral default intensity would be 8%. That is,

$$45.24 = e^{-(0.02+0.08)\times 1} \times 50$$

29.5.3 Relating the Credit Spread to Default Intensity and the Recovery Rate

Finally, let's consider the case where under default there is some recovery of face value. This amount is represented by $K \times RR$, where RR is the recovery rate (and the recovered amount is paid at time T). In this case, the value of the bond is given by:

$$\begin{aligned} D_0 &= e^{-r\times T}(RR \times K \times (1 - e^{-\lambda\times T}) + K \times e^{-\lambda\times T}) \\ &\approx e^{-(r+\lambda(1-RR))\times T} \times K \end{aligned} \tag{29.21}$$

When there is some recovery in case of default, the default intensity is reduced by the factor related to the recovery rate. The higher the recovery rate, the lower the impact of a default. The term $\lambda(1 - RR)$ in the second line of Equation 29.21 could be interpreted as the credit spread that is added to the risk-free rate to obtain the appropriate discount rate:

$$\lambda \times (1 - RR) \approx \text{Credit Spread} \tag{29.22}$$

The credit spread is smaller for bonds with higher recovery rates or lower default intensity.

APPLICATION 29.5.3

Suppose the credit spread on a one-year zero-coupon bond is 2%, and the recovery rate is estimated to be 80%. Find the risk-neutral default intensity.

Using Equation 29.22, the implied risk-neutral default intensity is 0.10:

$$\begin{aligned} 0.02 &= \lambda \times (1 - 0.8) \\ \lambda &= 10\% \end{aligned}$$

Note that Equation 29.22 contains three variables, any one of which can be solved using the other two.

29.5.4 The Two Predominant Reduced-Form Credit Models

Two models are most commonly cited when referring to the reduced-form models: the Jarrow-Turnbull (1995) model and the Duffie-Singleton (2003) model. The Jarrow-Turnbull model assumes that regardless of timing of default, recovery is received at the maturity date. The Jarrow-Lando-Turnbull (1997) model extends the original model further by taking into account various credit ratings beyond just two simple states of default or survival. Here, the model takes into account that there is migration risk (i.e., that the bond will be downgraded rather than experiencing outright default). The probability of moving from one rating to the next can be obtained from rating transition tables, published by the credit rating agencies. The Duffie-Singleton model allows the recovery process to occur at any time and sets the recovery amount to be a fraction of the nondefaulting bond price at the time of default.

29.6 PROS AND CONS OF STRUCTURAL AND REDUCED-FORM MODELS

Structural models have the attractive feature of connecting credit risk to the structural factors for the underlying borrowers, such as leverage (relative size of the debt compared to asset value) and valuation of the company. This connection is intuitive, and the theory provides an endogenous explanation of default. Furthermore, the adaptation of a well-accepted and fairly robust option pricing model to measure expected default is economically reasonable. Structural models allow for the valuation of debt while explicitly taking into account the capital structure of the borrower.

However, structural models make a few assumptions that are violated in the real world. One of the most challenging implementation issues has to do with firm values. The Merton model assumes that the value of the firm's assets is knowable. Although for a publicly traded firm one can observe the market value of the firm's equity, one of the largest real-life challenges is obtaining the value of the firm's assets. Another challenge in implementation is that the assumption of continuous tradability for corporate debt is not realistic given liquidity constraints in many sections of the credit market. Even for very liquid investment-grade corporate debt, there is a cost to trading and especially trading in size. Calibration of a stochastic asset process using public data is mathematically quite complex.

The KMV model further specifies different methodologies to calculate firm value. However, the chosen methodology depends on the availability of publicly traded equity to measure a firm's value. Naturally, the original challenges of inferring firm values remain for private firms. Another criticism of the KMV model is that the expected default frequency curve assumes that the return of the assets is normally distributed. Furthermore, this approach does not take into account the full complexity of the debt mix, such as seniority features, collateral and covenant packages, and special features such as convertibility. Last, it assumes a static capital structure. That is, once the debt is placed, the borrower cannot change the terms or restructure the debt. In practice, firms often have the ability to negotiate with their lenders—for example, to get a forbearance or waiver on an interest payment to avoid default.

Unlike structural models, reduced-form models assume that default (and loss given default) is exogenously determined. On one hand, it means that the model is more flexible and is easier to implement and calibrate. This is no small matter; in a real-life trading environment, a model that can quickly compute pricing based on rapidly changing observed market variables is both useful and user friendly. On the other hand, this model does not give the user an insight into the link between the fundamental drivers of credit-worthiness (leverage, capital structure) and default. In practice, structural models are typically more suited to fundamental security analysis or portfolio credit risk management, whereas reduced-form models are typically seen in fast-moving environments, such as the pricing of credit derivatives by credit trading floors.

29.7 EMPIRICAL CREDIT MODELS

Structural and reduced-form models attempt to model the economic and statistical underpinnings of bankruptcy and default. Structural models focus on the behavior of asset values relative to the amount of debt through time. Reduced-form models focus on the statistical process that generates default signals. Both models use these different approaches to obtain estimates of default probability and credit spread. The empirical models differ from these models in two important ways. First, empirical models believe that the default process is too complex to be modeled mathematically. Therefore, the focus is on using historical data on default to understand credit risk in a rather crude way. Second, empirical models do not attempt to generate an estimate of the probability of default or credit spread, at least not directly. The primary goal of empirical models is to create a credit score. A **credit score** is a measure that can be used to rank or assess the relative riskiness of firms or securities. The absolute values of these credit scores usually do not contain much useful information but rather are used on a relative basis to rank firms or securities in terms of their credit risk. For this reason, empirical models are sometimes referred to as credit scoring models.

29.7.1 The Purpose of Altman's Z-Score Model

In this section, we provide a brief overview of one such credit scoring model: Altman's Z-score (Altman 1968; Saunders and Allen 2010). First, we discuss certain characteristics of a business, to specify and quantify the variables that are effective indicators and predictors of corporate distress. Second, we discuss how a set of financial and economic ratios can be analyzed in a context of corporate distress prediction by creating a credit score for the business.

The **Z-score model** focuses on a set of financial ratios that are based on a firm's financial statements as well as the market value of the firm's equity. The ratios focus on those characteristics of firms that have proven to be useful in predicting financial distress, taking into account liquidity, profitability, leverage, solvency, and activity of a firm. Ed Altman, who created the Z-score in the 1960s, used a linear econometric model to determine how important each characteristic is in predicting financial distress. The following equation displays the econometric model estimated by Altman:

$$Z = 1.2 \times X_1 + 1.4 \times X_2 + 3.3 \times X_3 + 0.5 \times X_4 + 1.0 \times X_5 \qquad (29.23)$$

In this expression, Z is the resulting Z-score, or the credit score.

29.7.2 The Five Determinants of Altman's Z-Scores

There are five variables that determine Altman's Z-scores as depicted in Equation 29.23:

1. X_1: Working Capital/Total Assets. This ratio, frequently found in studies of corporate problems, is a measure of the net liquid assets of the firm relative to the total capitalization. Working capital is defined as the difference between current assets and current liabilities. Liquidity and size characteristics are explicitly considered. Ordinarily, a firm experiencing consistent operating losses will have shrinking current assets in relation to total assets.
2. X_2: Retained Earnings/Total Assets. This ratio measures the relative size of the total amount of reinvested earnings and/or losses of a firm over its entire life. This is the measure of the cumulative profitability of the firm over time and is

likely to be low for young firms. This ratio is indirectly related to leverage, since those firms with relatively high ratios are more likely to have financed their assets through retention of profits and not to have used as much debt.

3. X_3: Earnings before Interest and Taxes/Total Assets. This ratio is a measure of the true productivity of the firm's assets, independent of any tax or leverage factors. Since a firm's ultimate existence is based on the earning power of its assets, this ratio appears to be particularly appropriate for studies dealing with corporate failure. Furthermore, insolvency in a bankrupt sense occurs when the total liabilities exceed a fair valuation of the firm's assets, with the value determined by the earning power of the assets.
4. X_4: Market Value of Equity/Book Value of Total Liabilities. This measure shows how much the firm's assets can decline in value (measured by market value of equity plus debt) before the liabilities exceed the assets and the firm becomes insolvent. For example, the value of X_4 will be equal to 2 for a firm with a market value of its equity equal to \$1,000 and a debt of \$500, and the firm could experience a two-thirds [2/(2 + 1)] drop in asset value before insolvency. However, the same firm with \$500 equity will be insolvent if assets drop by only one-half in value.
5. X_5: Sales/Total Assets. This ratio, which is known as the asset-turnover ratio, is a standard financial ratio illustrating the sales-generating ability of the firm's assets. It is one measure of management's capacity in dealing with competitive conditions.

Altman used these five variables and historical data on a sample of firms—of which some defaulted and some survived—to estimate the coefficients that appear in Equation 29.23. The goal was to find a score with the maximum predictability power of default by the firms in the sample.

29.7.3 Solving for the Z-Score in Altman's Credit Scoring Model

Consider the information for PQR Corporation in Exhibit 29.1.

EXHIBIT 29.1 Financial Information of PQR Corporation

Assets	€	Liabilities and Equity	€
Current Assets	10,000,000	Current Liabilities	5,000,000
Fixed Assets	20,000,000	Long-Term Liabilities	10,000,000
		Retained Earnings	3,000,000
		Common Stocks	12,000,000
Total Assets	**30,000,000**	**Total Liabilities and Equity**	**30,000,000**
Sales	60,000,000	Market Value of Equity	18,000,000
Cost and Expenses	55,000,000		
Income before Taxes and Interest	5,000,000		

Given the definitions of X_1 through X_5, their values can be calculated as shown:

$$X_1 = 5{,}000{,}000/30{,}000{,}000 = 0.167$$
$$X_2 = 3{,}000{,}000/30{,}000{,}000 = 0.1$$
$$X_3 = 5{,}000{,}000/30{,}000{,}000 = 0.167$$
$$X_4 = 18{,}000{,}000/(5{,}000{,}000 + 10{,}000{,}000) = 1.2$$
$$X_5 = 60{,}000{,}000/30{,}000{,}000 = 2$$

Finally, the Z-score can be calculated as 3.49 using these five values and the coefficients in Equation 29.23.

29.7.4 Interpreting Z-Scores in Altman's Credit Scoring Model

While the absolute values of Z-scores do not have intuitive interpretations, one can use them to rank firms in terms of their levels of credit risk or likelihood of default. Using historical data on the performance of firms for which the Z-scores were calculated, Altman has developed the following rule for interpreting the absolute values of these scores:

- $Z \leq 1.81$: Default group
- $1.81 \leq Z \leq 2.99$: Gray zone
- $Z > 2.99$: Nondefault group

The 3.49 Z-score for the sample firm (PQR Corporation) indicates that this firm belongs to the nondefault group.

29.8 DISTRESSED DEBT INVESTMENT STRATEGY

Historically, the opportunity set in distressed investing came about because of the stigma associated with bankruptcy. For example, in the 1980s, many pension plans had rules that prohibited their portfolios from holding anything but Treasuries and investment-grade corporate bonds. As borrowers got into financial difficulties, the bonds had to be ejected from the portfolio, often at any cost. Over time, the investor community became much savvier about the bankruptcy process. At the moment, there is a set of investors with the right experience and skill set who understand that the bankruptcy and restructuring process can be a key step in bringing much-needed operational or financial fixes to a struggling borrower.[4]

29.8.1 Definition of Distressed Debt

When a company is in financial distress, that typically means that it is highly likely to fail or has already failed to meet its obligations to its lenders. There are two forms of default. The first is **technical default**—this is when a borrower is in breach of the covenants of its bond or loan obligations. The second is **actual default**, which is the failure of the firm to make the payment of interest or principal on its obligations as required in the contract.

The term *distressed debt* refers to the obligation of a company that is facing such financial difficulties. *Distressed* refers to the general condition of the borrower (financial distress) rather than a specific type of debt. A distressed debt fund can invest in a variety of instruments, such as leveraged loans, high-yield bonds, mezzanine bonds, or trade claims. Similarly, a distressed borrower, although typically referring to a corporation, can also refer to sovereigns, municipalities, or individual consumers. For example, the city of Detroit's municipal bond was referred to as distressed in the months surrounding its bankruptcy filing.

29.8.2 Trade Claims

Trade claims make up a specific group of creditors to a distressed borrower, such as the debt owed to a vendor that provides products or services to a firm. In a typical distressed situation, bank loan holders/lenders and bondholders typically form the majority of debt capital outstanding. Nonetheless, trade claims are an important constituent of the bankruptcy process. A company in distress is likely to face difficulties making payments to its suppliers for goods and services received. The vendor is unlikely to be willing or able to hold to its claim through the restructuring and exit, which can last for years. Thus, the vendor is likely to sell its claim, typically at a significant discount to the par value of the claim.

Note that, unlike bonds and similar to bank loans, trade claims are not considered securities, and therefore securities laws do not apply. The trading for these instruments tends to be less transparent and less liquid, adding to the potential ways a knowledgeable investor can add value. For the purpose of our discussion, distressed debt will refer to all debt instruments of a distressed borrower, including bonds, bank loans, and trade claims.

29.8.3 Size of Universe and Characteristics of Distressed Debt

There is no legal or formal definition for distressed debt. However, there are several market conventions as to how to measure the size of the distressed debt universe. First is by rating, such as by including debt instruments that have CCC or lower ratings. This measure includes instruments that the rating agencies expect to default shortly. Another measure by rating would include only instruments that are rated D or below. This is naturally more restrictive, as it excludes the instruments for which default is imminent and that are often already trading as if default is a given. This brings us to another measurement, which is measuring the distressed universe by the market price of the instruments. In the industry, the convention of a distressed corporate bond is a bond that is trading with a spread in excess of 1,000 basis points compared to a Treasury bond of comparable maturity. Another industry convention is to quote the price of a distressed corporate bond by price instead of by spread. A bank loan trading at a price of 80 (80% of par value) and a bond trading at a price of 40 (40% of par value) are often said to be distressed. Note that for all these ways of measuring the size of the distressed market, the convention is to include the amount of the bond at par. In other words, if a high-yield bond is

$100 million at issuance (at par) and the price of the bond has fallen to $40 million, one would use the par amount of $100 million to calculate the size of the distressed market.

Another indicator of a company in financial distress is the price of the equity. For borrowers that have their debt rated CCC and see their bond trading at a significant discount to par, the equity price is typically zero or very close to zero. Given market-based conventions as to what constitutes a distressed security, signs of distress are often well telegraphed in the price of the security well before they are reflected in the rating or the borrower files for bankruptcy.

Rating agencies view distressed exchanges and exchange offers as default. They either include these events in the overall default calculation or classify them as selective default (SD) or restrictive default (RD). A **distressed exchange** is classified by rating agencies as a default situation that occurs when the borrower repurchases the creditor's claim for cash, often at a great discount to the original par amount of the claim. An **exchange offer** is classified by rating agencies as a default that occurs when the borrower swaps out the original obligation for a new instrument, and likely for a reduced principal. In an exchange offer, the new instrument is likely to have a later maturity than the original paper and may or may not carry a higher interest rate.

29.8.4 Causes of Financial Distress

Distress can be caused by two main factors: lack of liquidity or lack of solvency. Lack of liquidity means that although the company is solvent, it lacks ready access to cash or other liquid assets. Lack of solvency refers to a situation in which the company's liabilities exceed the value of its assets.

Generally, a distressed investor considers lack of solvency to be a more serious problem than lack of liquidity. This is because the liquidity problem is temporary and can be fixed, typically by injection of either new debt or new equity. Both are readily available options if the underlying business of the borrower and the value of its assets are considered to be in excess of existing liabilities. However, lack of solvency usually stems from deeper problems, which are rooted either in the business or in its existing capital structure. Business-related issues can include cost structure, industry competition, or declining revenue. Issues with capital structure typically center on too much debt. A classic cause for financial distress is high debt levels paired with a cyclical business. Business-related problems typically require an operational restructuring, which tends to be more difficult and more costly to fix. An **operational restructuring** includes change in management, asset sale, spin-off, or various other ways of implementing cost reduction, including closing down or selling business lines. Problems surrounding too much indebtedness are typically resolved through financial restructuring to drastically reduce debt. For the majority of restructurings, both operational and financial fixes are often involved.

An investor in distressed debt is often said to look for a "good company, bad balance sheet." This means that even for a company that is insolvent, if insolvency is caused by too much debt heaped onto what is essentially a healthy enterprise with a clear reason to exist as an ongoing entity, then there is opportunity for the right set of investors. In this case, the focus of debt restructuring is typically to reduce the existing amount of debt so that the company can emerge from bankruptcy with a manageable debt burden. A debt restructuring often involves lengthy and extensive negotiation among different groups of stakeholders (creditors, management, and other equity holders).

In order to reduce the debt burden on the borrower, some creditors may give up all or a portion of their claims. There are a few reasons why lenders may do this. The creditor may be compensated with an equity stake in the newly restructured entity, or may be coerced to give up part of its claim as a way to avert greater losses.

29.8.5 Countercyclical Nature of Distressed Opportunities

The size of the distressed debt market varies significantly over time. One of the most attractive features of investing in distressed debt is the countercyclical nature of the opportunities. The opportunity set is typically richest when there is widespread default, and least attractive when default is scarce. A long-term investor can benefit from having an allocation to distressed debt as a diversifier in his portfolio. In addition, because of the existence of non-economic sellers, distressed investors can often purchase debt at very attractive prices that more than compensate for the significant default, liquidity, and process risk associated with distressed investing.

Long-term investors in the credit market have long observed that default is highly cyclical in nature. In other words, there is an ebb and flow to the economic cycle and, consequently, to the credit cycle. As the economic cycle expands, it is typically followed by increased availability of credit. This usually leads to more debt on the borrowers' balance sheets, often in combination with weaker covenants. At times, the thirst for yield exists for a prolonged amount of time, during which creditors are increasingly willing to accept less protection and lower yield from issuers. Unavoidably, there is a point where the cycle turns. When economic growth slows down, investor risk appetite changes, and new credit creation slows down dramatically or is shut down altogether. The default rate would then rise, along with the number of distressed opportunities.

Note that deterioration in fundamentals alone does not necessarily precipitate a default. A period of healthy economic growth usually overlaps a bullish credit market. In this environment, a company with declining business prospects can avert a default for a significant period. Investors' perception of the company's default risk can diverge significantly from the actual performance of the business, and this bullishness is reflected in sustained periods during which the company's bond trades at a yield similar to those of competitors with significantly better prospects. Furthermore, in a heated credit market, borrowers can issue new debt or "term out" their debt. **Terming out debt** takes place when borrowers extend the maturity of their existing debt. These actions greatly improve the liquidity available to the company and can mask operational issues long before they are reflected in the pricing of the debt.

29.8.6 Types of Investors and Investment Vehicles in Distressed Debt

The range of types of investors in the distressed debt space has grown tremendously over the past few decades, with the largest growth coming from institutional investors. This trend is consistent with that of increasing institutional involvement in alternative investment strategies, such as hedge funds and private equity. Institutional investors—such as private and public pension plans, as well as sovereign wealth funds—have a very long investment horizon. These investors and their long-term investment horizon are suitable for the time horizon that may be needed for some strategies used when investing in distressed securities. Furthermore, these investors need to balance return generation over the long term with capital preservation, and one way to protect against a big loss of capital in the event of a market downturn is to diversify by investing in strategies that profit from market downturns. Therefore, having part of their portfolio invested in distressed opportunities, with their countercyclical nature, is a sensible portfolio construction methodology for many institutional investors.

In terms of the types of vehicles used to invest in distressed securities, most investors in distressed situations hold their investments in either hedge funds or private equity vehicles. The debts of a company in financial distress have several characteristics that render them particularly suitable for investments through vehicles with flexible mandates.

The reasons for this are twofold. First, a hedge fund vehicle, unlike a traditional credit mandate, is not bound by ratings restrictions. As a particular bond gets into distress and rating agency downgrades follow suit, many investors with traditional credit mandates will no longer be able to hold the security. They will then have to sell the bond, often at the same time. The bond is then oversold, and the price falls dramatically, implying a yield that exceeds the expected loss in the event of default. Hedge funds specializing in distressed debt investing can then come in and benefit from their ability to be the liquidity provider to the investors with traditional credit mandates.

Second, investing in distressed securities often also involves capturing a liquidity risk premium in addition to capturing a default risk premium. This is particularly true for issuers or issuances that are smaller. In the event of a very large default, some of the defaulted securities can be very liquid. The debts issued by the largest entities in the Lehman Brothers bankruptcy are traded frequently; there is a large amount of issuance as well as numerous buyers and sellers for the various debts. However, good liquidity tends to be the exception rather than the rule when it comes to distressed investing. Part of the premium captured by investors comes from the ability to hold the investment during part of or throughout the restructuring process, when there are limited investors who are willing and able to take on the risk of owning the debt of a borrower in trouble. As such, the typical liquidity profile of distressed securities does not fit the mandate of investors who require their investment vehicle to have high liquidity. A case in point is that much of the recent growth in credit investing has come in the form of exchange-traded funds (ETFs) or other daily liquidity vehicles. The liquidity requirement of these types of investment vehicles is not appropriate for investing in distressed debt.

In an appropriately structured hedge fund vehicle, investors can have the flexibility to structure the liquidity profile of the investment vehicles to meet the underlying liquidity of the securities or the investment strategy. For example, some of the liquidity management features of a hedge fund vehicle include the ability to use side pockets. Credit hedge fund investors who experienced the market meltdown of 2008 are likely to be familiar with the notion of side pockets, which are effectively another share class of a fund in which the shares have no redemption rights.

Side pockets have received their fair share of criticism, and indeed in some cases they have been used inappropriately. Side pockets are inappropriate when used to house underperforming investments, such as previously performing debt for which the borrower has gotten into trouble so that liquidity dries up and the broader investment community shies away from the investment. In this case, side pockets enable a hedge fund manager to continue to collect performance fees from the rest of the (performing) portfolio without acknowledging the losses from the underperforming assets.

However, side pockets by themselves are neither good nor bad. An appropriate use of a side pocket is when it is used to house lower-liquidity or limited-edition investment opportunities (e.g., distressed debt). If the fund manager purchases a distressed debt issue and puts it in the side pocket at the time of inception (rather than moving a performing asset into a side pocket as soon as it becomes nonperforming), the side pocket provides existing fund investors with ways to manage their liquidity, particularly if the portfolio is largely made up of other credit investments with a better liquidity profile. If the investor needs liquidity, it can be provided from the rest of the portfolio, leaving the opportunity in the side pocket undisturbed while the restructuring takes place. Also, in cases in which there are multiple investors in the fund, if a significant part of the capital were to be redeemed, remaining investors would be protected from the risk of getting stuck with a disproportionate share of illiquid opportunities. In the opposite scenario—in which the opportunity set is limited, the distressed debt has seen large positive returns, and new investors come into the fund—existing investors' exposure to the distressed debt is protected, as the new capital does not dilute the existing investors' share in the distressed investment.

29.8.7 Return Drivers: Fundamental Valuation and Informational Inefficiency

In its classic form, distressed investing is the ultimate form of value investing. Value investing—that is, investing in situations that are out of favor—tends to be most profitable in a market that's ripe with informational inefficiencies. When it comes to the market for distressed securities, there are a few unique characteristics that explain why informational inefficiencies are particularly high. Investors with the ability to take on the fundamental credit risk of a borrower in trouble can be handsomely rewarded for their effort. However, there are significant difficulties in valuing a troubled borrower when the value of the business and its assets may be quickly deteriorating.

One of the reasons for the informational inefficiency is the siloed nature of the investor base in credit. A significant slice of the capital allocated to the credit market is invested under fairly constrained guidelines. A very common restriction on traditional credit investment vehicles involves ratings. Many investment guidelines specifically prohibit or severely limit the manager's ability to own debt that is rated CCC or below. Consequently, as a company runs into financial distress and its debt gets downgraded by the rating agencies, traditional investors would have to sell their holdings. At this point, the debt is often oversold, and the price falls dramatically, giving rise to opportunities for investors with unconstrained mandates, such as credit-oriented hedge funds.

Second, when a company falls into financial distress, the degree of complexity related to analyzing the future valuations of the company exponentially increases. Analyzing the prospects and options of a company in financial distress involves different skill sets from performing credit analysis on a healthy borrower. The typical fundamental credit analysis involving a healthy, creditworthy borrower would primarily look at the borrower's ability to service the cash flow and asset coverage. However, for a company that is facing declining business, running out of cash, losing customers at a fast clip, or facing quickly deteriorating asset value (or all of the above!), the traditional credit metrics and fundamental credit analysis do not apply.

To make matters worse, as a borrower goes into default, the sell-side coverage tends to disappear. If the borrower had publicly traded equities, as default approaches and the equity value falls to near zero, the equity is likely to get delisted. Delisting would result in reduction of sell-side coverage as well as removal of the public reporting requirement. In other words, as the borrower's financial situation goes from bad to worse, transparency into what's going on within the company, as well as informational efficiency, falls dramatically.

This means that there is great information asymmetry among different investors in the credit market as well as between creditors, management, and equity holders. Investing in the debt of a distressed company can be fraught with risk, and there are many investors who either are mandate constrained or lack the appropriate skill set to invest in this space. Thus, there can be significant mispricing in the debt of a distressed company, and the mispricing tends to be more pronounced in certain pockets, such as small and midsize issuers as well as off-the-run and non-index names. The flexibility of the hedge fund structure as well as the skill set of distressed debt hedge funds can give them the ability to separate the good value from the value traps, and the flexibility to use shorts and hedges further adds levers that distressed hedge funds can pull to add value for their investors.

29.8.8 Return Drivers: Event/Catalyst and Activism

Another return driver in distressed investing is the ability to benefit from positive events and catalysts. To a certain extent, distressed investing is a very specific subset of event-driven investing (default or restructuring being the specific event). As such, the strategy shares some similar return drivers with event-driven investing. Depending on the degree of active involvement the investor is willing and able to do, part of the return in a distressed investment can come from activism, or "making your own event." Although the term *activism* is sometimes associated purely with hostile strategies, there is a wide range of approaches in practice, including friendly activism. Here, the term will include a cooperative approach in which part of the investor's return comes from working with management and/or other stakeholders in improving value creation. The following will discuss some particular investment strategies in distressed investing to illustrate and highlight these drivers.

In a typical restructuring scenario in which there are different groups of creditors, an investor can add value by being actively involved in the restructuring, effectively creating her own event. In a chapter 11 bankruptcy, an investor can provide leadership by being part of the **steering committee** or the **creditors' committee**, committees in which investors have taken a cooperative approach to work among themselves and, in many cases, with management. Bondholders can be part of the official creditors' committee, which is typically made up of the seven largest unsecured creditors. A steering, or ad hoc, committee is usually made up of holders of similar debt claims working together.

Whether in steering or official creditors' committees, the goal is to propose a reorganization plan that maximizes the value of the group's interest. For example, a steering committee may propose a reorganization plan that includes securing new financing for the reorganized entity in exchange for the committee members getting part of their claims in cash and swapping the rest of their claims for new equity. In this scenario, the steering committee's ability to secure new financing is a positive event for the company. If the proposed plan is accepted by the bankruptcy court and the newly reorganized company survives as an ongoing entity, this is also a positive sign for the equity valuation, which can be an important return driver for the investor. In some cases, it can also be a very positive internal rate of return (IRR) event for management, which may receive some equity in the newly reorganized entity as part of its incentive package.

Note that part of the committee's duty is to evaluate the borrower's business and prospects, which will expose the committee members to material nonpublic information. This information is likely to be very valuable, especially given the information vacuum that often surrounds a distressed borrower. However, when an investor elects to be part of the official creditors' committee, the committee member is likely to face trading restrictions related to access to inside information.

In addition, being part of an official creditors' committee tends to be very time and labor intensive; for this very reason, many distressed debt investors refrain from being part of the official creditors' committee.

Another way investors can add value is to bring operational expertise that will make the business more attractive or more efficient. This typically happens in a situation in which the borrower is in need of operational turnaround in addition to financial restructuring. For example, if a borrower gets into distress because it has a few business units that are unprofitable, a distressed investor can work with management on selling the unprofitable units, focusing on the core competency of the business and making the whole enterprise leaner. This is another way that investors can drive returns by investors "making their own event" and bringing unique expertise to work collaboratively with management to unlock value.

29.9 BANKRUPTCY LAWS ACROSS THE GLOBE

When a company gets into financial distress, it has two options in restructuring its obligations. The first is to do an out-of-court restructuring and the second is to file for bankruptcy (chapter 11 in the U.S. Bankruptcy Code). An **out-of-court restructuring**, also known as a consensual restructuring, occurs when the borrower works with its creditors to reduce the debt burden. This type of restructuring, which tends to be more cost- and time-efficient, typically involves the exchange of an existing bond for a new bond for a smaller par amount and longer maturity date. The new debt may also have more flexibility for the borrower in managing its cash needs, such as an accordion feature. An **accordion feature** provides a borrower with enhanced flexibility, such as allowing the borrower to access a certain amount of additional borrowing, or allowing the borrower to have more access to liquidity, which is a matter of particular importance during distress.

A consensual approach is typically preferred, as a prolonged battle among stakeholders is costly and can even be harmful to the borrower's ability to remain as a going concern. For example, when General Motors (GM) was in bankruptcy, unless it was made clear that GM would emerge from bankruptcy, it would have been hard to convince customers to buy cars and for suppliers to continue to provide goods and services to a car company that was about to disappear.

However, a consensual approach is not always possible or realistic. The more complex the company's capital structure is, the more stakeholders there are, which makes an out-of-court negotiation nearly impossible. In this case, the restructuring is likely to be done in court. The next sections discuss some of the bankruptcy regimes across the globe.

29.9.1 Bankruptcy in the United States

In the United States, a bankruptcy filing typically involves two sections in the U.S. Bankruptcy Code. Chapter 11 governs restructuring, and chapter 7 governs liquidation. Typically, most bankruptcy filings are followed by a restructuring. If an attempt to restructure the company as a going concern under chapter 11 fails, then the process may move to a liquidation of assets under chapter 7. Effectively, the U.S. Bankruptcy Code provides a clear procedure for valuation of the debtor's assets and ranking of competing claims.

Bankruptcy can be filed voluntarily by the debtor or involuntarily by its creditors. A bankruptcy filing, whether voluntary or otherwise, rarely comes as a surprise to the market. A filing can come well after a long and arduous attempt to restructure out of court failed and the relationship between the debtor and the creditors soured. Note that in practice an involuntary filing is rare. The borrower is usually aware when it is about to run out of liquidity and likely to fail to meet an interest or principal payment, and typically files voluntarily. By filing voluntarily for bankruptcy, a borrower becomes known as a **debtor-in-possession** (**DIP**) and maintains control over the in-court restructuring process.

A chapter 11 filing can come as a **prepackaged filing** ("prepack"), which means that when the debtor files for chapter 11, it does so with a plan of reorganization (POR) that has already been distributed, solicited, negotiated, and voted on by all the relevant classes of creditors. Having the relevant stakeholders come to an agreement before coming to court usually translates into a very short restructuring process, and one that is very cost-efficient.

A prolonged time spent in bankruptcy court is expensive. There are bankruptcy attorneys, trustees, experts, bankruptcy advisers, turnaround consultants, and a slew of other service providers who are involved. The expertise of these workout/restructuring specialists is costly, and those payments are higher in seniority than payment to creditors. In addition, if the company were to continue operating as a going concern during the workout period, it would likely need liquidity to finance working capital. Capital providers will require higher yield and collateralization to supply financing to a company that is currently in bankruptcy. Consequently, a prepack chapter 11 filing is preferable to a "free-fall" filing—that is, a bankruptcy filing without a POR.

A chapter 11 filing provides some structural benefits for the debtor-in-possession (i.e., management). First, it will have the exclusive rights for a period after filing to propose a POR and even more time after filing to solicit votes for its POR. Under the 2005 Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA), the exclusivity period is limited to 18 months, and the solicitation period is limited to 20 months.

If management's POR fails to get enough votes for confirmation, different creditor groups are eligible to submit their own PORs. Different classes of claims eligible to vote would have to vote on a POR to be confirmed for exit from chapter 11. Grouping claims into classes is an important part of the U.S. Bankruptcy Code. First, it determines voting eligibility. Second, after payments due to administrative and priority claims, the U.S. Bankruptcy Code ranks priority of payment by the type of class. From the most to least senior, these are: secured claims, priority unsecured claims, general unsecured claims, and equity claims.

It is important to note that the bankruptcy law groups claims that are "substantially similar." There is no clear guideline as to what the phrase means, and in practice, similar claims may or may not be classified together. When it comes to borrowers with multiple business units, subsidiaries, and creditors, the different intercreditor agreements, various liens, and multiple intercorporate guarantees add complexity to the seniority of different claims and the assets that can be used to satisfy those claims.

An important feature of U.S. bankruptcy law is what's referred to in the industry as a cramdown, which refers to the ability of a bankruptcy court to involuntarily impose a POR on a class of creditors that had previously voted against that POR. In other words, a POR and exit out of chapter 11 is possible without having to have the POR approved by all classes of creditors or equity interests. In effect, this gives the court the ability to arbitrate among competing interests, ending what could be an endless and value-depleting fight between various stakeholders. As long as the plan meets the relevant provisions, the U.S. Bankruptcy Code allows for the plan to be confirmed even with objections as long as the plan is "fair and equitable" and does not "discriminate unfairly" against each rejecting class.

Generally, the "fair and equitable" test is dependent on whether the rejecting class has secured claims, unsecured claims, or equity interests. A plan is considered to discriminate unfairly against a specific class if another class that has equal or more junior ranking is to receive greater value under the plan without a reasonable justification for that treatment.

29.9.2 Bankruptcy in the United Kingdom

England and Wales are typically considered to be a creditor-friendly environment. Many restructurings go through a popular court-sanctioned process called **scheme of arrangement**, which can be used to bring about a reorganization of a solvent company or group structure, as well as an insolvent restructuring, including a debt-for-equity swap or other debt-reduction methodologies. In practice, schemes are often used to avoid insolvency. For a scheme to be approved, it needs a minimum 75% approval as measured by value of each class of creditor and majority by number in the impacted classes. In this process, cramdown of minority votes within a class is permitted. If a scheme is approved, then the court will decide on whether the scheme is sanctioned, including reviewing the fairness of the approved scheme. With a court's sanction, the scheme would then become binding on the company, its members, and creditors.

Another option is through **administration of restructuring**, which transfers the control of the debtor into the hands of an administrator. This process is designed to conduct workout and restructuring of an insolvent company; in practice, however, it often leads to the liquidation of the business and its assets. Prepackaged restructuring is an option; that is, the debtor enters the administration process with a preapproved plan. Note that contemplation of a company sale requires that the valuation of assets be done on a "best price achievable" basis. In other words, the sale process (i.e., liquidation) will determine the price and thus the creditors' recovery. This is in contrast with the U.S. view, in which valuation in chapter 11 is based on an "objective" or a "fundamental" basis—meaning that a comparable sale or other methods of valuation can be used. In practice, the U.S. view tends to be more supportive of keeping the debtor as a going concern rather than to go through liquidation/sale. Other European countries apply a concept of valuation that is similar to the UK view.

Some issues remain to be addressed; these include how to provide funding to an insolvent entity in the middle of restructuring (i.e., unlike in the U.S. courts, there are no clear rules on the mechanism for DIP loans), as well as how to resolve various issues of intercreditor claims. Generally, however, the UK government has shown a lot of progress in moving the country's bankruptcy laws closer to U.S. chapter 11–style rules in order to save more companies from liquidation.

29.9.3 Bankruptcy in the Rest of Europe

Investing in distressed opportunities outside the United States and the United Kingdom comes with a significant amount of jurisdiction risk. Even in areas where there are fairly well defined bankruptcy laws, such as in continental Europe, the difference in insolvency procedures across jurisdictions can translate into very different outcomes for creditors. There have been movements afoot to develop a more consistent procedural process across Europe. In the past few years, some German, Italian, and Spanish companies have shifted jurisdiction to the United Kingdom to access the scheme of arrangement, and UK companies have used the scheme to restructure both their English-law-governed and their non-English-law-governed liabilities.

GERMANY. In 2012, the German parliament enacted a broad package of bankruptcy reforms that would align Germany's restructuring laws with practices in the United States and the United Kingdom. The fundamental changes in the reform created new opportunities to preserve the entity as a going concern. Furthermore, they created a process of potential transfer of equity to creditors in a court-supervised process, something that is virtually impossible in the existing legal regime. In the new rules, creditors will have more say in the selection of an administrator and provide a DIP-type loan. However, note that old habits die hard. Generally, a consensual out-of-court approach is still the preferable method unless the debtor is obliged to file for insolvency.

FRANCE. Generally, the regime here is regarded as neither creditor friendly nor debtor friendly but rather social friendly (for the social good), which means that the interest of another party—typically the workers—takes priority. Generally, there are two types of pre-insolvency processes (typically consensual). One has no time limit (*mandat ad hoc*); the other, *conciliation*, can take up to five months. During these processes, a *mandataire* or a *conciliateur* is appointed by the court to facilitate discussions between interested parties. There are three types of insolvency proceedings. The first is the *sauvegarde* (safeguard), and it is typically debtor led. The debtor is responsible for presenting a plan, there are two creditors' committees (financial and suppliers), and two-thirds in value of each committee plus a general bondholder meeting is required to vote for the plan. Under the *sauvegarde*, the judge can force creditors to a 10-year debt extension, and security cannot be enforced. The second is the *reglement judiciaire*, which is broadly similar to the *sauvegarde*, but the court has more control over the borrower and is able to conduct asset sales. The third is the *liquidation judiciaire*, which is effectively a liquidation of assets, conducted by a court-appointed liquidator. However, there have been recent changes in France that bring the rules closer to the U.S.-based chapter 11. A new process called *sauvegarde financière accélérée* (accelerated financial safeguard) allows for a proactive consensual debt restructuring and is a pre-insolvency, court-sanctioned process. In this process, the creditors' committee is able to present its own plan, and the process can impose binding results on minority holdouts.

29.9.4 Bankruptcy in the Rest of the World

For the rest of the world, bankruptcy laws are generally not well developed. In Asia, the Middle East, and Russia, value realization for creditors often lies outside the legal system. In areas where equity ownership of many corporate borrowers is concentrated in the hands of a few powerful family interests, going to court to enforce creditors' rights is not likely to be fruitful or even yield any resolution. In these cases, the option for an investor is either to invest in private equity or to focus on the larger borrowers that depend on continued access to the global debt market. (It is important to note that a distressed debt investor can invest in the debt issued by a borrower outside the United States or Europe but only under U.S. or English law, which governs most bank debt trading across Europe.)

29.10 IMPLEMENTATION OF DISTRESSED DEBT STRATEGIES

This section discusses four methods for implementing distressed debt strategies.

29.10.1 Implementation through the Loan-to-Own Control-Oriented Approach

One of the better-known strategies in distressed investing is the loan-to-own approach. **Loan-to-own** is a strategy that is typically seen as being on the more hostile spectrum of approaches and begins with the investor providing a loan that can lead to the investor taking control. In the event that the borrower gets into financial distress and is unable to meet its obligations to the lender, the lender can decide to come in, restructure their debt claim, and swap their debt claim for equity and control of the business. In order to effectively execute the strategy, the lender needs to either be the sole lender or have a controlling share of the debt. Liquidity in this strategy is low, the restructuring process can be very short or can take years, and the fund manager typically has the ability to hold on to the equity long after the debt-for-equity swap takes place. Returns in this scenario come from two main sources: the coupon payment of the original loan before distress takes place and the appreciation of the value of the equity after the debt-for-equity swap takes place. As part of the debt-for-equity swap, the principal of the original loan is likely to be written down in part or in full. The split of the return between the coupon payment and the equity value depends on how long the debt is performing before distress sets in, as well as the depth of the distress and the time it takes to restructure the business.

29.10.2 Implementation through the Classic Approach

The **classic distressed investing strategy** involves buying the loan or bond of a distressed borrower, holding the instrument through workout/bankruptcy, swapping part or all of the debt claims for equity of the newly reorganized entity, and finally exiting via sale to a strategic buyer or the public market through an initial public offering (IPO). Typically in this strategy, the investor specifically seeks to invest in the fulcrum security of the debtor's capital structure, which is the security that is most likely to be converted from debt to equity as part of the workout or restructuring process. Although an investor may not be the only lender, typically an element of control or at least a way to influence the process is a key part of the investment thesis.

Loan-to-own is not the only distressed investing strategy in which an investor can exert control. As previously discussed, an investor can exert control by being part of the official creditors' committee or an ad hoc steering committee. Another control-oriented approach is to buy large amounts of the pre-petition bonds or loans. Depending on the size of the portfolio he manages, a fund manager can buy enough size to be influential, all the way to owning the majority or supermajority position of the outstanding bonds. Naturally, the latter is easier to do for smaller issuers and issuances, and is very difficult to do for very large restructurings. One thing to note is that investors need to be aware that having control or ability to steer the restructuring to a favorable outcome is not necessarily tied to holding enough bonds to own the majority or the supermajority of votes. Although the official creditors' committee is typically made up of the largest bondholders, there is no restriction for other (smaller) investors who wish to form an ad hoc committee.

There are several ways in which investors with significantly less than a majority holding can still influence the votes. First, if some decisions require a supermajority vote (e.g., two-thirds), then an investor holding a third of the eligible vote plus one effectively has veto right (i.e., although this investor cannot force the rest of the group to vote for a particular plan, he can block plans deemed to be unfavorable to his position). Second, an investor with the right skill set and leadership ability can influence others in the group of investors. By doing the analysis showing the other creditors and the bankruptcy judge how her plan is better than the other PORs, an investor with a smaller holding can influence the outcome of the restructuring. Third, investors of similar mind-set can group together and form their own group, eventually proposing their own POR.

29.10.3 Implementation through the Trading-Oriented Approach

There is a myriad of strategies that investors can use to generate attractive returns in distressed situations, including non-control-oriented strategies. Typically, a trading mind-set is seen for non-control-oriented strategies. The key differences here include the length of the holding period for the investment, the size of the investment (absolute and relative to the total issuance size), the liquidity profile of the instruments, the return expectation, and the use of hedges.

In a **trading-oriented distressed strategy**, the goal is to capture the excess return premium that's implied in the price of debt instruments that are oversold when non-economic sellers have to liquidate their holdings. At the beginning, when distress starts to set in, prices may fall well below even the most dire recovery scenario. A nimble investor, spotting the opportunity, can go in early and then exit when the price starts to recover from the troughs after more investors (likely with a longer-term horizon) come in following a much lengthier and more involved due diligence. Unlike a loan-to-own or a classic distressed investor, a trading-oriented investor is likely to trade in and out of the position throughout the period before filing, during the restructuring, and after the borrower exits the bankruptcy process. Unlike other types of investors, a trading-oriented investor tends to get into the instruments that are more liquid to facilitate trading in and out, and is unlikely to try to influence the restructuring or the negotiation. A trading-oriented strategy is more suitable for large restructurings where there are many buyers and sellers for some or all of the issuances. A large restructuring also makes it possible for a trading-oriented investor to own only a small percentage of the market, thus avoiding the impact of buying and selling on price. A trading-oriented investor also tends to have a lower internal rate of return (IRR) target than the control-oriented or classic distressed investors. This is because the latter groups assume more risk—they hold the position for a longer term, typically take on greater liquidity risk, and risk being caught in a restructuring that lingers for years as the underlying business continues to deteriorate.

Last, a trading-oriented investor may also employ hedges to further manage the risks. Classic distressed and loan-to-own investors are mostly long-only or very long-biased, as distress is typically a highly idiosyncratic risk and difficult to hedge. However, investors owning the defaulted bonds of Lehman Brothers may realize that, given the high rate of mortgage exposure in the Lehman Brothers estate, the prices of those defaulted bonds may vary with the situation in the mortgage market. Consequently, these investors may hedge their long bonds with a mortgage index to reduce their general exposure to the broader mortgage market and isolate their target risk on improved valuation of Lehman bonds.

29.10.4 Implementation through a DIP Loan Approach

If a distressed borrower were to stay as a going concern and maintain operations during its period of restructuring, then it would need liquidity to finance its operations. Capital providers will require higher yield and collateralization to provide financing to a company that is currently in bankruptcy. For a company in chapter 11, liquidity can be provided via debtor-in-possession (DIP) loans, which typically earn a higher coupon (and can be issued at a discount to par) than a bank loan of similar maturity and enjoy a super seniority status compared to the indebtedness incurred before the bankruptcy filing. This means that if the restructuring fails and the assets of the borrower need to be liquidated, the DIP loan has to be paid fully and in cash before the proceeds can be distributed to other creditors. If the restructuring succeeds and there is a POR that is supported by the creditors, the DIP loan has to be paid in full and in cash before the POR can be confirmed and the borrower can exit chapter 11.

An investor in a DIP loan gets to reap additional yield because it is acting as a liquidity provider to a company that is currently in distress. The super seniority, however, usually means that the loan is well covered by the asset value and, combined with the higher yield, makes a DIP loan a very attractive proposition for many lenders. In the past default cycle, there have been cases in which there was more supply than demand for DIP loans. In many situations, the priority to provide DIP loans goes to the existing pre-petition lenders. In situations in which new lenders come in to provide the DIP loan, the loan typically carries a higher coupon rate than if the DIP loan came from existing lenders. This reflects two potential situations: One is that a new lender that is less familiar with the situation at a company in distress would assume the worst and need to be compensated accordingly to provide capital. A second possibility is that there may be actual deterioration in the true health of the borrower, where existing lenders are unwilling or unable to provide more liquidity. Similarly, a new lender would need to be compensated accordingly.

29.11 VALUATION RISKS IN DISTRESSED DEBT INVESTING

When it comes to assessing the value of a distressed entity, an investor would have to make some assumptions on the potential future of the business. A distressed borrower is akin to a sick person; a small bump in day-to-day operations (such as a small increase in the number of days for accounts receivable) that would not have a material impact on a healthy business can quickly translate into a crisis when liquidity is scarce and management is distracted by the restructuring rather than focusing on running the business.

Distressed investors need to be aware that current financial metrics such as revenue, cost, and profitability can deteriorate quickly, and thus projections need to include a fairly wide band of downside probabilities. For example, it is common that when a borrower goes into distress, its suppliers are no longer willing to extend any credit and will demand to be paid cash on delivery instead. This increases the working capital needs for a borrower that is already strapped for cash. Thus, it is key that distressed investors consider the value of the business as an ongoing entity as well as the liquidation value of the assets.

29.11.1 Liquidity Risk in Distressed Investing

If the lack of buyers for distressed opportunities provides one side of the investment coin, then the drop in liquidity is the other side of that coin. There may be a long period during which there is very little trading in the distressed security. This drop in liquidity may be reflected in the wide bid-ask spread of the instruments, if any trading happens at all. For example, under regular market conditions, a “distressed” bank loan may command a bid-ask spread in the range of 1 to 3 points, while a distressed bond commands a bid-ask spread of 2 to 4 points. However, in the time of heightened market volatility, it is not uncommon to see a distressed bank loan command a bid-ask spread of up to 10 points, or 20 points for a distressed bond. This means that if an investor who has entered into a distressed situation then decides to exit the position, there would be a significant cost to do so. At the height of distress, most of the nondistressed investors have typically left the arena, and the ones who are left are likely to be professional distressed investors. Their requirement for a substantial risk premium is usually reflected in the bids for the position should an investor need to sell.

29.11.2 Mark-to-Market Risk in Distressed Investing

The thinner liquidity for distressed securities also gives rise to significant volatility in their prices. Often, the prices of securities may show a big drop in a vacuum of new information. If there is a seller who needs to sell part or all of his holdings, there may not be many buyers for the security, and the sale may occur at a significant discount to the last recorded market price. Investors with a sizable amount of holdings in distressed securities need to be aware of this and be able to withstand significant mark-to-market volatility in the portfolio. Because of this, a price return or performance-based stop loss would not be appropriate and would likely crystallize losses in a distressed debt portfolio.

29.11.3 Legal and Jurisdiction Risk in Distressed Investing

One of the key skill sets needed to invest in distressed debt is the ability to understand how the bankruptcy laws and rules vary across the globe. Additionally, beyond a deep understanding of the law, a successful distressed investor needs to possess a good understanding of the day-to-day operations of the court and the past rulings of specific bankruptcy judges. Because of this, many distressed investors tend to specialize in specific jurisdictions in order to reduce the risk of getting involved in situations in which they don't feel that they sufficiently understand the finer points of creditors' rights and how the relevant bankruptcy court(s) treat creditors' rights. For example, the bankruptcy laws in the United States and the United Kingdom tend to be considered the most creditor friendly, whereas the laws in Italy and France are considered less creditor friendly.

Even when investing in the United States, investors in distressed securities need to be aware of the risk of so-called forum shopping. In the event of a voluntary bankruptcy, the debtor may file in a particular bankruptcy court that it expects will rule the most favorably toward the debtor or current management. In this scenario, having familiarity with past rulings of a particular judge or a particular court would be of great value for a distressed investor. An example of how forum shopping can add significant complexity and risk to the situation is found in the case of Caesars Entertainment. In this case, one group of bondholders filed for involuntary bankruptcy in one bankruptcy court, only to discover that the debtor had filed for voluntary bankruptcy in another bankruptcy court. The latter bankruptcy court had been known to rule in favor of the debtor in similar situations. There is now an additional layer of complexity, as it needs to be decided which court will prevail, as it will impact the treatment of bondholders and the value they are expected to recover in the restructuring process.

29.12 ASSET-BASED LENDING

An **asset-based loan** (**ABL**) is a secured loan backed by various types of collateral pledged by the borrower. The value, type, and quality of the collateral determine the amount of the loan that can be extended, the advance rate, and the interest of the loan. Hedge funds and private equity funds making these loans are said to be part of the shadow banking system, as traditional lenders may reduce credit availability when capital requirements are increased by regulators worldwide.

29.12.1 A Typical Borrower in Asset-Based Lending

The typical beneficiary of an asset-based loan is a small or midsize company, and the typical size of the ABL ranges from $10 million to $50 million, although there are loans that are significantly larger. Borrowers can be in various industries and sectors, including retail, distribution, manufacturing, wholesale, and service companies.

A typical ABL candidate has an asset-rich balance sheet, with a significant amount (half or more) of its total assets in working capital. These assets are generally in the form of inventory or accounts receivable. Similar to other types of lending, an asset-based lender looks for a borrower with a strong management team that has a proven track record of managing the operational complexities of its business, strong if not market leadership in its industry, and a strong ability to service its debt. Furthermore, given the nature of the ABL, it is important that borrowers have strong financial accounting and information technology (IT) systems that enable reliable data on asset performance as well as on operational results.

Unlike in the cash flow–based lending market, the ABL market relies far less on external credit ratings (i.e., those provided by credit rating agencies such as Moody's, Standard & Poor's, and Fitch). Given the collateralized, secured nature of asset-based lending, this type of lending is typically done with borrowers with credit risk similar to non-investment-grade borrowers (i.e., borrowers with a rating of BB+ and below for Standard & Poor's or Ba1 and below for Moody's).

Geographically, the ABL market is well developed in the United States, and there are established markets in Europe and Canada as well. Lenders have extended credit to multinational companies with assets residing in geographies outside the areas just mentioned, with certain caveats. Lenders need to be able to get security interest in the underlying collateral, as well as gain comfort that, in the event of restructuring, there is precedent on how creditors can get recovery on their loan. Practically speaking, structures such as a special purpose vehicle to own the collateral as well as a **lockbox** (a bank account set up to receive collections of the accounts receivable—proceeds on recent activity to be applied against a revolver) have been used to protect the lender.

29.12.2 Why Borrowers Select Asset-Based Lending

In the corporate credit arena, a lender can lend to a corporation based on two different frameworks: asset-based lending or cash flow–based lending. In the cash flow–based lending model, a cash flow–based lender provides secured loans and typically looks at earnings before interest, taxes, depreciation, and amortization (EBITDA) adjusted for cash as a proxy for cash flow, combined with a multiplier to determine how much leverage a company can take. During business and economic downturns, many businesses will see their absolute and relative earnings contract. This often translates into difficulty accessing financing. To make matters worse, during difficult economic conditions, lenders typically also become more bearish and adjust their leverage multiple downward, further shrinking the amount of credit available to a particular borrower.

For an asset-based lender, the loan is secured by particular assets, and many asset-based lenders feel that asset valuation tends to be less volatile than earnings multiples when it comes to determining available credit to a borrower. This also benefits certain types of borrowers because the availability of credit tends to remain more stable over different business and economic environments.

29.12.3 Features of Asset-Based Lending

ABL lenders use a key metric called borrowing base. **Borrowing base** is used in deciding how much credit to extend to a particular borrower and is determined by the amount of eligible collateral a borrower can pledge. In addition, the type of eligible collateral and the mix of assets within the collateral package determine an advance rate. The **advance rate** is the ratio of credit for every dollar of collateral. For example, for an asset with an advance rate of 80%, if the asset is valued at $100, $80 of credit can be extended against it. Generally speaking, assets that are easier to value and have better liquidity are deemed to be more attractive as collateral by the lenders, who tend to apply a higher advance rate in calculating the borrowing base.

Accounts receivable collateral is measured by various factors, such as the diversification of client base (the more diversified, the better), length of term, typical collection period, and dilution in determining the advance rate. In most cases, advance rates on typical portfolios of accounts receivable range from 75% to 85%. In some cases, it makes sense for the borrower to provide credit insurance and boost the advance rate of the accounts receivable to 90%.

The specific terms involving valuation of the underlying collateral as well as the advance rate are highly negotiable and vary dramatically across borrowers. For example, the advance rate against a particular borrowing base can fluctuate over time. **Seasonal overadvance** is a temporary allowance by which the lender allows for a higher advance rate to account for seasonal effects in which the working capital need of the borrower is higher. For example, retailers' busiest season is typically in November and December, and in preparation for that season, they usually need to build inventory. The need to purchase more goods will lead to higher working capital need, and a seasonal overadvance provides some flexibility to allow higher borrowing.

Traditional overadvance is typically used for a corporate transaction, such as an acquisition or a leveraged buyout (LBO). In a **traditional overadvance**, lenders allow for a greater advance rate, and the additional borrowing is then amortized over several years and can be added to an existing term loan or as a separate facility. In connection to the amount of borrowing in excess of the regular advance rate, there may be an excess cash flow sweep (i.e., additional cash flow is redirected to repay the additional borrowing) to bring the loan back to the regular advance rate. The metrics used to analyze an overadvance are closer to traditional cash flow lending inasmuch as the lender looks at expected ability to service the debt, historical performance of the company, and the total leverage of the borrower.

In evaluating the value of inventories, the lender typically hires a third party who is a professional appraiser. The appraiser will usually have past experience in handling the liquidation of the particular type of inventory. Measures such as how quickly the inventory can be converted into sellable merchandise, profit margin, whether the inventory is a commodity, and the mix of the inventory all factor into the appraisal value and ultimately into the borrowing base. Next, an advance rate is applied to the borrowing base to determine the amount of the loan. The advance rates seen for inventory range pretty widely depending on the type of inventory and the expected ease of liquidation. Common industry practice is to set the value of the inventory as a percentage (80%–90%) of liquidation value, net of the cost of liquidation. For commodity-oriented inventory, the advance rate can go as high as 80%.

29.12.4 Use of Asset-Based Lending Proceeds

The use of an ABL typically varies depending on the type of borrower. Larger, higher-quality borrowers typically use asset-based loans as a way to finance working capital. Because of their size and credit quality, these are typically the borrowers that can finance most of their liabilities through public or private credit markets. These borrowers may use ABLs either to fund seasonal changes in their working capital (i.e., borrow to finance higher inventory during restocking, and the inventory collateralizes the borrowing) or to finance certain corporate actions, such as opportunistic asset purchases, share repurchases, or special dividends.

Smaller and midsize borrowers tend to use ABLs more extensively in their overall capital structure. For these borrowers, the asset-based borrowing may include longer-term loans that are backed by hard assets, such as real estate or machinery. Additionally, companies who are in active acquisition mode may also use ABLs as part of their acquisition financing in addition to other, more junior debt.

In addition, borrowers that are in stress or distress may also rely on ABLs, since their credit profiles render them unsuitable for cash flow–based lending. These types of borrowers typically need capital that they can rely on as management restructures the business or weathers the current situation. In this scenario, the ABL may act as a bridge loan that provides some temporary liquidity as well as operational and financial flexibility as the owners execute a turnaround. As the turnaround is completed, the company may repay the ABL and go back to financing its liabilities in the cash flow market.

29.12.5 Asset-Based Loan Structures and Collateral

A typical ABL credit facility is composed of a revolver and a term loan. A **revolver**, or a **revolving line of credit**, is a credit line with a preapproved limit that's available for a prespecified period. As the borrower takes out some capital, the amount of available credit is reduced by that amount. Unlike a term loan, the borrower can repay the amount of borrowing at any time before the end of the period, and the same amount will again become available for future borrowings. The borrower typically uses the revolver to finance working capital needs. As the need for working capital fluctuates over time, the revolver structure allows the borrower to access more credit as the business needs grow and to repay the loan (and free up future capacity) as cash comes in or business slows down. A **term loan** typically has either an amortizing or a bullet structure and is secured against longer-term assets (real estate, machinery). Often, the term loan is provided against capital expenditures such as new machinery. Generally speaking, the term loan typically makes up a third or less of the total ABL facility. Typical tenors for ABLs range from three to five years; however, the terms of these deals tend to be highly customizable and can involve significant negotiation between the borrower and the lender.

The most common assets to back an ABL facility are inventory and accounts receivable. In other words, these are assets with fairly good liquidity and easy-to-determine value. As discussed, longer-term assets (real estate, equipment) tend to be used to back term loans. An important part of the negotiation is determining what types of assets are not eligible for inclusion in the collateral package of the ABL. For example, past-due account receivables and incomplete work-in-process inventory are commonly excluded from the collateral package.

As the ABL market matures, lenders have grown to acknowledge that borrowers may be able to offer valuable collateral in the form of intangible assets, such as brand name and intellectual property (e.g., patents and trademarks). As expected, valuation of intangible assets tends to be more complex than that for tangible assets such as inventories. Third parties with particular expertise in evaluating the value of specific intangible assets are likely to be involved, as well as legal counsel (also likely with specific expertise, such as expertise in intellectual property law), to make sure that the relevant asset comes with a perfected security interest. Market practice has shown that borrowers that are able to borrow against intangible assets tend to be more creditworthy borrowers whose brand name is well established in the market and/or that have intellectual property that can be sold or monetized separately from the physical products of the company.

29.12.6 Asset-Based Lender Protection and Covenants

For a cash flow–based loan, one of the key covenants is the **net leverage covenant**, which is calculated as the amount of senior or total debt, net of cash, as a multiple of EBITDA. However, for an asset-based loan, there is typically no leverage covenant. This means that the ABL borrower has more flexibility in the amount of borrowing that it can incur, especially if the business has a volatile earnings stream.

ABL lenders are more focused on ensuring that the borrower has enough liquidity in order to meet its cash needs. Liquidity can be measured as either available cash or unused capacity under the existing facility given the borrowing base. Therefore, an ABL lender is focused on analyzing the cash flow from the underlying operations of the business relative to cash needs, such as interest payments on the loan (and other debts, if they exist), taxes, any dividends or distributions to shareholders, and capital expenditures. The one credit metric that ABL lenders tend to be focused on is the fixed charge coverage covenant. The **fixed charge coverage** is a ratio equal to (EBIT + fixed charge)/(fixed charge + interest). The key to this calculation is determining what constitutes a fixed charge. Common fixed charges include rent/lease, utilities, insurance, and salary.

In practice, some asset-based facilities are structured such that the fixed charge coverage covenant does not come into play until the borrower's unused loan capacity falls below a certain limit. This gives the borrower additional flexibility, as the liquidity limit does not kick in until a certain point at which the lender indicates that the borrower's access to liquidity has fallen below a certain level (typically 10% to 20% of the asset-based facility). Naturally, larger and more creditworthy corporate borrowers tend to have better access to this type of feature, also known as a springing covenant.

Another feature for some asset-based facilities is the lender's ability to block the borrower from accessing a portion of the loan if the borrower fails to meet the liquidity test. Although this seems to make asset-based loans less attractive to borrowers, the borrower benefits from not having to worry about meeting financial covenant tests, and having to default on the loan if it breaches a covenant. For asset-based loans, a typical covenant that limits debt level relative to enterprise value or cash flow is seen as less relevant. From the borrower's perspective, there is a trade-off between flexibility and access to capital.

In addition, certain features of asset-based facilities can provide substantial flexibility so that the borrower can take corporate actions. For cash flow–based loans, there are often negative covenants. These nonfinancial covenants effectively prohibit the borrowers from undertaking certain actions that are seen as detrimental to the lenders. An example of a negative covenant is the prohibition on borrowers paying a special dividend to the shareholders of the company. Without this type of negative covenant, a borrower such as a family-owned business is able to distribute cash to the owners as long as the liquidity test is met.

In practice, lenders also closely monitor the changes in the collateral base. In addition to the basic checks of ensuring that the collateral actually exists, the changes in the amount and the quality of the collateral are also closely monitored, typically on a daily basis. This enables the lenders to closely track the day-to-day status of the assets backing their loan and to catch any potential problem before the borrower gets into a severe liquidity crunch.

29.12.7 Asset-Based Lending Risk

Asset-based lending can involve specialized or unique risks, including the following five.

29.12.7.1 Valuation Risk The nature of the underlying assets backing the loan tends to change on a day-to-day basis. Accounts receivable change daily as new accounts are added to the mix and existing accounts are converted into cash. Inventories change as finished goods are sold and new materials are purchased. Similarly, the mix of inventory changes as raw materials enter production and work in process converts parts into finished products, sometimes within the span of a day. This change in the assets and, consequently, the value at which these assets can be liquidated is not always reflected in periodic financial statements.

Thus, asset-based lenders bear the risk that the value of the collateral backing the loan may be different from what they expected. Lenders mitigate this risk by closely monitoring the collateral. Periodic reports on inventory composition, third-party valuation, and accounts receivable aging are examples of how lenders manage valuation risk. These periodic reports are provided by the borrower in frequencies ranging from monthly to weekly or even daily. The most creditworthy borrowers tend to be able to report inventory composition or accounts receivable aging on a monthly basis, but in extreme scenarios lenders may require daily reporting. Third-party valuation reports tend to be costly, and, in terms of balancing risk mitigation and cost, these reports may be done on a semiannual or quarterly basis rather than more frequently.

29.12.7.2 Risks Regarding Process and People Given the data and labor-intensive monitoring required on the underlying assets, lenders need to have a fairly sophisticated system that enables easy reporting by the borrowers as well as easy information retrieval for the lending team. In addition, systems alone are not sufficient. The lender needs a dedicated team with the right skill set to collect and analyze the data from the borrower. This is particularly true for borrowers that submit data daily. Monitoring of information and the ability to detect signposts for deteriorating liquidity, such as increasing number of days to collect accounts receivable, is key in mitigating losses. Therefore, if the lender does not have the right process and people, data collection alone is not entirely useful. In fact, having more data can backfire, as the systems are overwhelmed and the lender is not able to get useful information from the borrower data.

29.12.7.3 Risks Regarding Hedging Asset-based lending is generally a long-only strategy. Given the smaller nature of a typical borrower, the highly idiosyncratic nature of small businesses, and the unique features of the specific ABL facility, hedging is difficult. Hedging a long portfolio filled with small issuers with a short portfolio composed of larger bond issuers (because it is difficult to find borrowers for middle-market bonds) introduces basis risk, in which the lender is betting on the health of small businesses versus larger businesses. In historical data, smaller businesses tend to be more impacted by business cycle downturns, whereas larger businesses have more levers to pull, including more diversified business lines and more ways to access capital, such as existing relationships with banks or the ability to tap capital markets globally.

29.12.7.4 Legal Risks One of the risks in the event that the lender needs to collect the collateral is that it finds itself in a position where it is unable to exercise ownership of the collateral. In a secured transaction, there are very important steps to make sure that the lender has legal rights to take possession of collateral in the event of default; under U.S. laws, these steps are known as **attachment of security interest**. In addition, lenders need to **perfect the security interest**. The party that owns a secured security perfects the security interest in order to help assure that no other party, such as another creditor or a bankruptcy trustee, will be able to claim the same collateral in the event that the debtor becomes insolvent. There are various methods for perfecting a security. For example, legal documents can be prepared to perfect a security, or the party can take possession of the collateral. The process will be different for different jurisdictions. In addition to credit/default risks, there are also legal risks to the lender if the paperwork required to attach and perfect the security interest is not complete or is not done properly. Lenders need to have access to the right legal expertise and representation to make sure that they do not find themselves stripped of the collateral that was expected to be theirs in the event of default.

29.12.7.5 Risks Regarding Timing/Exit In the event that the borrower gets into distress and the lender needs to seize the collateral backing the loan, there is the risk that the collateral value has fallen below what the lender is owed. For example, a loan backed by inventory that is largely composed of seasonal merchandise, such as toys, may lose its value after the holiday season. Furthermore, if the merchandise does not get sold quickly and consumer tastes and preferences change, that would further harm the cash value of the collateral. Not all risk is removed just because the lender is able to take control of the collateral, as the time taken to liquidate the collateral and changing market conditions can translate into significant exit risk for the lender.

29.13 CONCLUSION

This chapter has introduced three different credit risk and credit scoring models. The structural and reduced-form models attempt to generate the same outputs—probability of default and credit spread—but they use different economic approaches. While the structural models focus on the borrower's balance sheet, the reduced-form models focus on modeling the random process that signals financial distress. The third approach discussed in this chapter uses data from financial statements of borrowers to create a credit score. This score may be used by lenders to decide whether to extend credit to a borrower and at what rate.

This chapter has also introduced two credit investment strategies employed by hedge funds: distressed debt and asset-based lending strategies. The crucial point about these two strategies is that they require significant amounts of highly specialized skills if their investors are to add substantial value. The fact that these strategies require such skills makes the market for the underlying instruments rather inefficient, which is perhaps the most important source of alpha for these strategies.

NOTES

1. For further discussion of adverse selection and moral hazard in financial markets, see Mishkin (2012).
2. For further discussion of credit risk models, see Duffie and Singleton (2003).
3. For further discussion of Merton, KMV, and reduced-form models, see Saunders and Allen (2010).
4. For further discussion of distressed debt investing and credit hedge funds, see Pascualy (2013) and Whitman and Diz (2009).

REFERENCES

Altman, E. 1968. "Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy," *Journal of Finance* 23 (4): 589–609.

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 8 (3): 637–54.

Crosbie, P., and J. Bohn. 2003. "Modeling Default Risk." White paper, Moody's KMV Company.

Duffie, D., and K. Singleton. 2003. *Credit Risk: Pricing, Measurement and Management*. Princeton, NJ: Princeton University Press.

Jarrow, R., D. Lando, and S. Turnbull. 1997. "A Markov Model for the Term Structure of Credit Spreads." *Review of Financial Studies* (Summer): 481–523.

Jarrow R., and S. M. Turnbull. 1995. "Pricing Derivatives on Financial Securities Subject to Credit Risk." *Journal of Finance* 50 (1): 53–85.

Merton, R. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29:449–70.

Mishkin, F. 2012. *The Economics of Money, Banking, and Financial Markets*. 10th ed. Upper Saddle River, NJ: Prentice Hall.

Pascualy, P. 2013. *Investing in Credit Hedge Funds: An In-Depth Guide to Building Your Portfolio and Profiting from the Credit Market*. New York: McGraw-Hill Education.

Saunders, A., and L. Allen. 2010. *Credit Risk Management in and out of the Financial Crisis: New Approaches to Value at Risk and Other Paradigms*. Hoboken, NJ: John Wiley & Sons.

Whitman, M., and F. Diz. 2009. *Distressed Investing: Principles and Technique*. Hoboken, NJ: John Wiley & Sons.

CHAPTER 30

Volatility, Correlation, and Dispersion Products and Strategies

Investors are familiar with common positions with returns that are tied to the directional returns of underlying real or financial assets. This chapter discusses products and strategies with returns that are tied to changes in levels of asset volatilities, correlations, or dispersions. These increasingly popular products and strategies can enable investors to better manage the risks of their portfolios and to better position their portfolios to benefit from superior market predictions.

30.1 VOLATILITY, RISK FACTORS, AND RISK PREMIUMS

Traditional risk premiums offer increased returns for increased exposure to systematic risk. For example, investors may seek to earn the equity risk premium, perhaps 4% or more per year over a riskless rate, by holding long positions in stocks or stock market indices. Investors also can own investment-grade and high-yield corporate bonds, both of which are expected to earn a credit risk premium of perhaps 0.5% and 1.5% net of defaults annually over the yield on low-risk sovereign debt, respectively, in compensation for lending to borrowers who have greater or lesser ability to repay the funds borrowed (Lorenz, Lin, and Tsang 2014). This section discusses returns from products with values that are tied directly to the levels of return volatilities.

30.1.1 Volatility as a Return Factor Exposure

Factor exposures are the underlying drivers of the returns to assets. As discussed in Chapter 2 of this book, recent advances in asset allocations have pointed out that the risk premium on an asset class represents a basket of risk premiums reflecting the various factor exposures of that asset class. These factor risk premiums vary over time, with positive returns earned when financial markets and the economy are performing well, and experiencing losses during economic slowdowns. This section introduces and discusses volatility as being one of the most important factors that impact a wide variety of asset classes.

For example, options prices are linked to the anticipated volatility of the assets underlying the options as seen by the direct relationship between option prices and volatility in option pricing models such as the Black-Scholes option pricing model. Thus, volatility is clearly a factor exposure to option returns. Of course, option returns also contain large factor exposures to the returns of the underlying assets. In recent years, a large and growing set of volatility derivatives has emerged. **Volatility derivatives** are pure plays on volatility with returns that are driven substantially, explicitly, and directly by exposure to the volatility factor.

30.1.2 The Volatility Factor Exhibits Negative Market Risk and a Negative Risk Premium

The level of return volatility exhibits a negative correlation to the returns of the market index. In other words, higher levels of volatility occur contemporaneously with market declines whereas lower levels of market volatility occur with market advances. Therefore, most traditional asset classes tend to have a negative exposure to the volatility factor. For instance, the returns to long-only equity and credit strategies decline in times of rising volatility. When an investment tends to decline with rises in return volatility levels, the position is said to be **short volatility**. While being short volatility is traditionally associated only with option writing, long positions in traditional asset classes such as equities are short volatility.

It is important to note that while we typically think of positive risk premiums to be attached to positive exposures to factors, investments with positive exposures to volatility factors carry a negative risk premium related to that factor. Conversely, everything else being equal, the greater the extent to which an asset has a negative exposure to the volatility factor, the higher the positive risk premium. This is evidenced by products that are pure plays on volatility. The Chicago Board Options Exchange (CBOE) Volatility Index (VIX) futures contract, to be detailed later in the chapter, is positively correlated to equity market volatility (i.e., is long volatility), and it therefore tends to decline in value due to a negative risk premium. An investor establishing a long position in the VIX contract enjoys the hedging benefit of a negative equity beta but pays for that protection through the negative risk premium. In other words, products that are long volatility have a **negative volatility risk premium**, which means they tend to lose money on average (relative to the riskless rate).

30.1.3 Volatility as an Unobservable but Unique Risk Factor

Volatility as a risk factor is not immediately observable; therefore we need to estimate it or find proxies for it. In addition, volatility is not directly traded and investment products must be created in order to trade volatility. A number of methods have been developed for presenting and measuring statistical properties of the volatility factor. The two most commonly used measures are implied volatility and realized volatility. Implied volatility is derived from the price of liquid options, whereas realized volatility is estimated by observing a time series of returns.

Implied volatility can be isolated and traded using options, and products exist where the implied volatility of an asset can be observed and traded. Since the financial crisis of 2008–9, as investor attention to understanding and trading volatility has increased substantially, the number of these products has increased, along with their trading volume.

As discussed in Chapter 2, legitimate risk factors are supposed to carry risk premiums that cannot be explained by other risk factors. True risk factors should be unique rather than highly correlated with each other. For example, it was shown that the value and momentum factors are unique since they are not highly correlated with each other or other factors (e.g., size factor or overall market factor).

In this context, supporting the claim that volatility is a unique factor, Rennison and Pedersen (2012) show that there is a low correlation (i.e., less than 30%) between the return to selling straddles and returns to the underlying markets. They also find that the correlations between commodity volatility versus commodity futures, currency volatility versus carry strategies, and interest rate swaptions versus long credit investments are also less than 30%. Equity volatility and long stock investments have a 50% correlation of returns. Volatility strategies across the four markets (commodities vs. equities vs. rates vs. currencies) also have low correlations of between 17% and 32% for all pairs. This evidence not only supports the idea that volatility is a risk factor, but also supports the idea than assets directly related to volatility are a unique asset class.

30.1.4 Using Volatility Derivatives to Hedge Market Risk

Most investors have substantial investments in traditional asset classes, and therefore are long market risk and short volatility. Some investors take long positions in volatility derivatives to hedge their traditional portfolios. That is, the gains from long volatility investments tend to offset the losses from their risky traditional asset portfolio during times of crisis and rising market volatility. As previously discussed, products and positions that are long volatility tend to lose money on average due to the negative risk premium of volatility factor exposures. There are two reasons for this.

First, as discussed in Chapter 2, risk factors that perform poorly during poor economic conditions must offer a positive risk premium. The positive risk premium is the incentive that investors need to hold a positive market beta asset that is expected to perform poorly when other risky assets are also expected to perform poorly. Therefore, since long volatility products are expected to perform well when other risky assets are performing poorly, they have a negative market beta and carry a negative risk premium.

Second, because most traditional asset classes are short volatility, natural demand for long volatility investments (for the purpose of hedging risk) exceeds the natural supply of these products. Therefore, market participants who supply long volatility products are expected to earn a positive return for taking a short volatility position. Investment products that allow investors to access this risk premium attached to the volatility factor are considered a separate asset class.

30.2 USING OPTIONS TO MANAGE PORTFOLIO VOLATILITY EXPOSURE AND RISK PREMIUMS

The previous discussion mentioned volatility derivatives as direct plays on volatility exposures. This section discusses option strategies that can be used to manage portfolio exposures to volatility risks and returns.

30.2.1 Option Writing as a Short Volatility Strategy

Many investors seeking to earn a risk premium through short volatility strategies typically sell short-dated options products. The positive risk premium of being short

EXHIBIT 30.1 Time Decay (Theta) Accelerates toward the Expiration Date of an Option

volatility can be seen in the time decay (theta) of short option positions. The theta of an option depends on the tenor of the option. The **tenor** of an option or other contract is the length of time until the contract terminates. Option investors can manage risk exposures to their portfolios by establishing positions in options with desired tenors and other characteristics.

An investor who desires to limit the risk of unhedged short volatility positions may choose to hedge this risk by buying long-dated options that experience time decay at a slower rate. Time decay of options moves at the square root of time, as illustrated in Exhibit 30.1. Consider a 12-month option that, by definition, has 100% of the premium of a one-year option. A six-month option, therefore, has 71% (square root of one-half) of the premium of a one-year option, while a three-month option has 50% (square root of one-quarter) of the premium of a one-year option (assuming each option has the same implied volatility). Because the time value (or theta) of an option decays at the square root of time, a long position of a one-year option held for nine months and a short position of a three-month option held until expiration have a similar cost of carry (i.e., profits or losses, holding all other variables constant). A single position in an equity call or put is not a pure play on volatility because it contains substantial directional risk to the underlying equity. Note that a portfolio consisting entirely of short positions in equity call options contains a substantial negative market beta whereas a portfolio consisting entirely of long positions in equity call options contains a substantial positive market beta. The next section discusses option straddles that tend to lower this directional risk.

30.2.2 Writing Option Straddles and Strangles as a Short Volatility Strategy

A long option straddle contains a long position in a call option and a long position in a put option on the same asset with the same expiration date and strike price. A strangle is analogous to a straddle except that the strike prices of the call and the put differ. A **short straddle** position contains a short call option and short put option on the same asset and with the same strike price. A **short strangle** position contains a short call option and short put option on the same asset but with different strike prices. Straddles and strangles were detailed and diagrammed in Chapter 6 of CAIA Level I.

Straddles can offer a more direct play on volatility than a position in a single call option or a single put option. To varying degrees based on moneyness and time horizon, the directional risks of the call and put in a straddle offset each other, therefore reducing the straddle's market beta. However, both the call and the put are short volatility, meaning that their vegas do not offset each other and therefore the straddle offers a purer play on volatility than a position in only a call or a put. A long straddle, like a long call or put, is long volatility whereas a short straddle is short volatility. An at-the-money (ATM) straddle has a near zero market beta and therefore offers exposure to the volatility factor with minimal short-term exposure to directional risk in the assets underlying the options.

A short straddle offers a relatively pure play as a short volatility position and is therefore poised to earn the short volatility premium on average. Decreasing volatility along with relatively small directional moves will allow the short straddle position to earn small positive returns through time decay. However, an increased level of directional volatility will tend to cause losses from either the short call or the short put, causing the short straddle to experience potentially enormous losses in highly volatile markets. In practice, sellers of straddles and strangles will dynamically hedge the delta risk of their options positions, seeking to limit the losses from directional moves. Delta hedging makes these positions closer to a pure play on volatility.

30.2.3 Writing Option Butterflies and Condors as a Short Volatility Strategy

While selling straddles or strangles offers a moderately pure way to earn the volatility risk premium, it exposes sellers to potentially unlimited losses should the underlying asset make a large move. Some investors prefer positions with more limited risk such as iron butterflies or iron condors, which are illustrated in Exhibit 30.2. In a short position in an **iron butterfly**, the trader sells a bull spread and a bear spread such that the two spreads share the same middle strike price. In a short position in an **iron condor**, a trader sells an out-of-the-money bull spread and an out-of-the-money bear spread. Bull spreads and bear spreads were detailed in CAIA Level I.

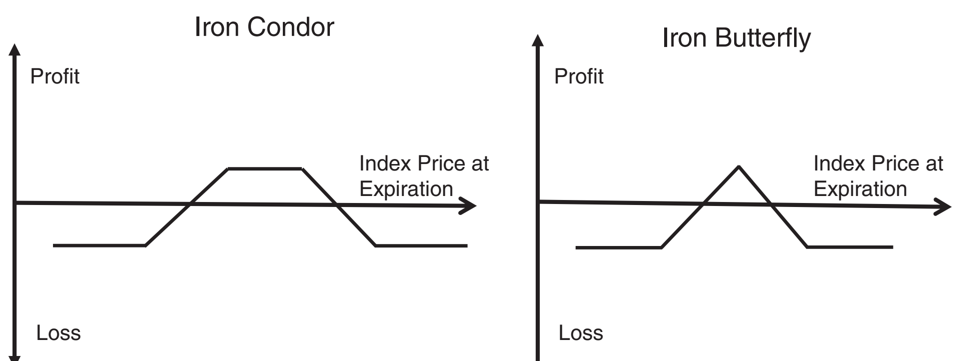


EXHIBIT 30.2 Payoff Diagrams for Long Iron Condor and Iron Butterfly Strategies

In both a condor and a butterfly, the tails of the risk exposures have limited exposures to large directional moves that are in contrast to the potentially unlimited tails of short straddles and strangles. Black and Szado (2016) show that from 1986 to 2015 short iron butterfly and iron condor positions on the S&P 500 Index earned over 2% annual positive alpha relative to the S&P 500, while experiencing beta risk of only 0.17 and 0.13, respectively, to the S&P 500.

30.2.4 Risk Management of Option Portfolios Using the Greeks

The risk exposures of various option strategies and option portfolios are often expressed using the Greeks—risk measures explained in CAIA Level I. Most investors in volatility strategies seek to create positions with predetermined levels of sensitivity to changes in implied variance (i.e., reach target exposures to vega) while minimizing exposure to small and large changes in price of the underlying asset (i.e., delta and gamma). The trader may seek to earn the volatility risk premium by implementing positions that short vega and theta. The rate of volatility and the length of time combine to form the volatility $\sigma\sqrt{T}$ over the tenor of an option. Vega and theta are therefore very highly correlated risk measures because they appear together throughout option pricing models within the product $\sigma\sqrt{T}$ (although they are not perfect substitutes, because T appears as a square root and T is also used in discounting cash flows). Nevertheless, vega directly reflects the effects of declines in volatility of the returns of the underlying asset whereas theta directly reflects the passage of time.

Vega and theta are themselves determined by the anticipated level of volatility. Therefore, changes in the level of implied volatility change the gamma and theta of an option. Options with higher levels of the implied volatility will have a lower level of gamma. That is, the option's delta sensitivity to the change of the underlying price is lower. Options with higher implied volatility will also have a higher level of theta. That is, holding all else equal, an option with a higher implied volatility will decay more each day than an option with a lower implied volatility.

An alternative to reducing delta risk through hedging calls and puts with each other is to hedge a call position with a short position in the underlying asset or to hedge a put position with a long position in the underlying asset. Some products, such as VIX futures and variance swaps, have exposure to vega exclusively without the need to participate in delta hedging trades, as is the case with portfolios of call or put options.

30.2.5 The Volatility Exposure of Delta-Neutral Option Portfolios

A position in one or more options that is hedged to be delta-neutral (i.e., hedged with respect to changes in the option's underlying asset) is a play on the volatility of the options' underlying asset. If the portfolio is long (short) options then it is long (short) volatility.

30.2.5.1 Long Volatility Exposure from Long Gamma The long volatility exposure of a long option position is indicated by its positive gamma; short options are short volatility and have negative gamma. A delta-neutral portfolio with positive gamma benefits from high volatility, suffers from low volatility, and is neutral with respect to the direction of the underlying asset.

Consider a long position in 100 at-the-money call options with a delta of 0.60. Assume that the options have a price of $4 and their underlying asset is a stock with a market price of $50. A delta-neutral hedge of 100 calls would require the short sale of 60 shares of the underlying asset (i.e., the 0.60 delta times the 100 options being hedged). Hypothetically, if the option's delta remained constant at 0.60 then a $10 change in the 60 shares of stock (for a total change of $600) would be exactly offset by a $6 per option change in the 100 options (also a $600 change, but in the opposite direction). But the option's positive gamma ensures that its delta will change such that there will be a net gain for the portfolio regardless of whether the stock price rises or falls (assuming that no time passes and the option's implied volatility does not change).

For example, if the stock rises by $10, the option's price will rise by more than $6 so that the gain from the 100 options will exceed the $600 loss on the short position in the underlying asset. If the stock falls by $10, the option price will fall less than $6 so that the loss on the 100 options will be smaller than the $600 gain on the short position in the stock. Therefore, the delta-hedged portfolio gains in value for an instantaneous change in the stock price in either direction. To keep a long gamma portfolio hedged, the trader buys as the stock falls and sells as the stock rises.

However, option prices generally decline through time, everything else equal, due to their theta exposure. Therefore, the profitability of the delta-neutral, long gamma portfolio is driven by two offsetting effects: the negative effect of the passage of time (theta) and the positive effect of volatility in the underlying asset (gamma).

30.2.5.2 Realized Volatility, Implied Volatility, and Profitability The performance of a delta-neutral portfolio of an option and its underlying asset is driven by realized volatility (i.e., the actual volatility of the underlying asset while the portfolio is being held). The net profitability of the hedge depends on the relationship between the implied volatility of the option when the portfolio was established (i.e., the initial implied volatility) and the subsequent realized volatility. Generally, if the realized volatility of the underlying asset exceeds the initial implied volatility of the option, then the positive effects of long gamma will dominate the effects of theta for a gain. Conversely, if realized volatility is low, the portfolio will tend to be more negatively affected by the passage of time than positively affected by volatility.

Since the initial implied volatility of the option reflects the level of future volatility that the market anticipates in the underlying asset, the success of an investment strategy depends on the ability of the investor to identify options with implied volatility that misestimates expected realized volatility. A delta-hedged long option portfolio will earn an attractive return if realized volatility exceeds the market's anticipated volatility (ignoring transactions costs and subject to assumptions regarding the underlying assets volatility).

30.2.5.3 Effects of Re-hedging on Portfolio Profitability A key dimension of delta-hedging is the frequency, if any, with which hedging activity is used to re-hedge the portfolio to delta neutrality after a change in the price of the underlying asset has moved the portfolio into being exposed to delta risk. The frequency of this hedging activity or rebalancing can have an important effect on performance depending on the type of volatility that occurs in the underlying asset. Specifically, the effect of rebalancing frequency depends on the type of volatility that occurs: trending (directional), mean-reverting, or neither (random-walk). For example, no rebalancing works best in trending markets. To illustrate, let's consider two cases of rebalancing a delta-neutral portfolio that is long the option (i.e., long gamma): infrequent rebalancing and frequent rebalancing.

Infrequent rebalancing: Without rebalancing, a delta-neutral, long gamma portfolio will benefit greatly from volatility that is directional but will have little or no benefit from volatility that is mean reverting. For example, if the underlying stock price rises dramatically but then falls back to its original level (i.e., mean reverts) before being rebalanced, the short equity position will return to its value but the long option position will decline in value due to its theta for a net loss. Conversely, if the stock trends up or down, the long gamma portfolio will earn a high return if the portfolio is not rebalanced. Thus, infrequent rebalancing is a bet that volatility will be directional (trending). A portfolio with infrequent rebalancing will experience poor performance when the underlying asset mean reverts even if the realized volatility in the short-term returns of the underlying asset is very high.

Frequent rebalancing: A delta neutral, long gamma portfolio that is frequently rebalanced to maintain delta neutrality will benefit moderately from any volatility including directional, mean-reverting, or random returns. The portfolio will earn an attractive return prior to transactions costs as long as the volatility of short term returns (e.g., daily returns) of the underlying assets exceeds the initial implied volatility.

30.2.5.4 The Role of Hedging Activities in Long Gamma, Delta-Neutral Portfolios

Hedging activity *per se* does not generate profits. Long gamma and volatility generate profits. Further, hedging activity (rebalancing) detracts from portfolio returns to the extent that the returns of the underlying asset are trending. However, hedging activity may recognize profits and may lower the risk of the strategy by returning the portfolio to delta neutrality. Note also that the hedging activity of frequent rebalancing can exacerbate drag from transactions costs.

The decision to be long or short options in a delta-neutral portfolio is a play on the level of realized volatility. The decision of how often to rebalance the portfolio to delta neutrality is a play on the tendency of the underlying asset to mean-revert or trend. When an asset is trending (mean-reverting), the realized and annualized volatility of its daily returns will be lower (higher) than the realized and annualized volatility of its weekly or monthly returns. Thus the comparison of realized and implied volatility depends on the computation interval (granularity) of returns.

The key points are fourfold: (1) Delta-neutral portfolios that are long options are long gamma (long volatility) with profitability that is driven by the relationship between the realized volatility of the underlying asset and the implied volatility of the option when the position is established. (2) The time interval selected for rebalancing a hedge back to delta neutrality will affect the profitability of the position depending on whether the underlying asset's price changes tend to be directional, mean-reverting or random. (3) Rebalancing the portfolio to maintain delta neutrality will enhance returns prior to transactions costs only if the returns of the option's underlying asset exhibit mean reversion. (4) Hedging activities (rebalancing) lower risk and accelerate the recognition of profits, but hedging activities lower performance when the returns of the underlying asset are trending and when transactions costs are included.

30.2.6 Six Properties of Realized Volatility

As detailed in CAIA Level I, the realized return volatility of an asset is the actual variation (typically measured as the standard deviation) that occurs over a specified time period using a specified return measurement interval. Obviously, understanding volatility products and strategies relies on understanding the properties and behavior of realized volatility. Sinclair (2013) presents six stylized observations regarding realized volatility, many of which are key assumptions behind some volatility arbitrage portfolio strategies and risk management techniques:

1. Realized volatility is not constant. Realized volatility mean reverts. Evidence also suggests that realized volatility clusters and has a long memory, which indicates that mean reversion is not very fast. As such, many traders will model volatility using a variety of generalized autoregressive conditional heteroscedasticity (GARCH) and regime switching approaches.
2. Realized volatility tends to stay low for some extended period of time until a market shock occurs and volatility transitions to a higher level for some period of time.
3. The volatility of short-term changes in realized volatility can be high, but in the long run, volatility tends to revert toward some long-term average level.
4. Empirical and laboratory evidence suggests that higher volatility increases investors' risk aversion. This indicates that higher realized volatility tends to be negatively correlated with returns on most risky assets and positively correlated with returns on most safe assets; see Smith and Whitelaw (2009) and Guiso, Sapienza, and Zingales (2013).
5. Equity market realized volatility tends to increase in bear markets and decrease in bull markets. In addition, a decline in the stock price of a firm increases its leverage and riskiness, resulting in higher volatility.
6. The rate at which realized equity volatility rises in bear markets exceeds the rate at which realized volatility falls in bull markets.

30.2.7 Implied Volatility Structures of Derivatives

Implied volatility is the estimate of anticipated return volatility over the remaining life of an option that is inferred from the current option price and a specified option pricing model (along with the model's inputs other than volatility). Although implied volatility is usually estimated using market prices, it does not generally represent an accurate and unbiased consensus of the realized volatility expected by market participants. In the case of the Black-Scholes option pricing model, implied volatility computations rely on values such as the price of the underlying asset and the riskless rate, which can be readily observed. The problem with the Black-Scholes option pricing model is that its assumptions are highly restrictive, including the assumption that the underlying asset has a constant volatility. Other option pricing models relax the assumption of a constant volatility but, in turn, can involve parameters requiring a great deal of subjective judgment.

Implied volatilities from options that differ by tenor, by moneyness, and by direction (i.e., put vs. call) tend to differ. Advanced option and volatility strategies often focus on the relationships between the implied volatilities of different options. An

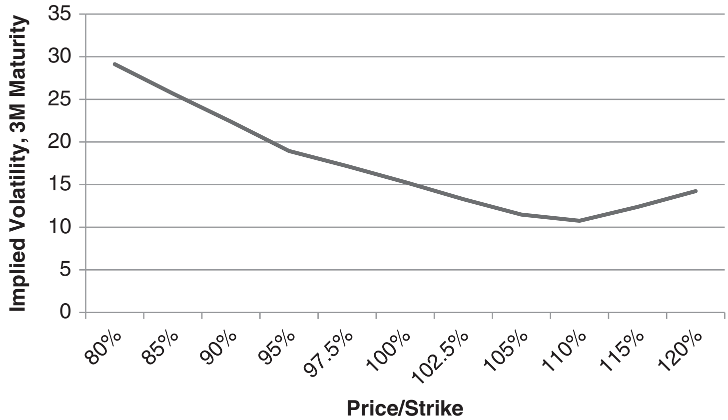


EXHIBIT 30.3 Example of Volatility Smile or Smirk from 30-Day S&P 500 Put Options on April 26, 2016
Source: Bloomberg.

implied volatility structure is a representation of the various implied volatilities relative to their tenor, moneyness, or type.

Exhibit 30.3 illustrates a popular implied volatility structure that focuses on the relationship between implied volatilities and moneyness. The exhibit illustrates a volatility skew. A **volatility skew** indicates that options that differ by moneyness have different implied volatilities.

Exhibit 30.3 shows out-of-the-money (OTM) put options, at-the-money (ATM) put options, and in-the-money (ITM) put options. Put options are ATM when the price/strike is near 100%, OTM below 100%, and ITM above 100%. Exhibit 30.3 indicates ITM options as having lower implied volatility than ATM options, which in turn have lower implied volatility than OTM options. In equity index options, the graph for a single expiration date seems to look like a smile or a smirk. A volatility structure with a **smile or a smirk** is where out-of-the-money put options have higher levels of implied volatility than other options. OTM call options tend to have a lower implied volatility than put options a similar distance from the current market price. An **options volatility surface** is a volatility structure that plots implied volatility for a wide variety of options in a given instrument across both expiration dates and strike prices. For example, an options volatility surface, as shown in Exhibit 30.4, would essentially combine the information from Exhibit 30.3 across all expiration dates for a given asset.

30.2.8 Evidence That Short Volatility Earns a Positive Risk Premium

Chapters 6 and 19 of CAIA Level I contain lengthy discussions of options and volatility arbitrage strategies. Rennison and Pedersen (2012) examine whether investors with short volatility positions are able to earn a consistent profit when implied volatility exceeds realized volatility—that is, whether options are frequently overpriced such that the volatility indicated by implied volatility exceeds the expected realized volatility. They estimate the volatility risk premium to be about 10% of the level of implied volatility, which can be earned through options and volatility trading strategies

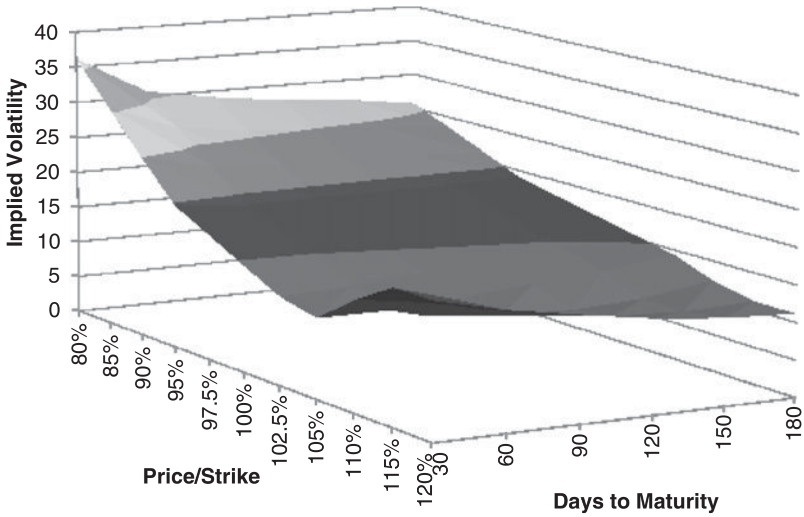


EXHIBIT 30.4 Example of Volatility Surface
Source: Bloomberg.

in the equity, interest rate, currency, and commodity markets. Exhibit 30.5 quantifies the returns to short straddle positions across 14 different options markets and details the claim that the risk-return trade-off for volatility strategies can compare favorably to the risk premiums that can be earned by investing in equity and credit products.

If being short volatility earns a positive risk premium, then being long volatility should tend to pay a risk premium, thereby generating low or negative average returns. The VIX futures contract (to be detailed later) and several VIX-related exchange-traded funds (ETFs) offer a pure play on exposure to the U.S. equity market volatility factor. The VXX exchange-traded fund demonstrates the long-term effect of positive volatility exposure using the VIX futures contract and therefore indicates a pure play on being long volatility. One of the most profound demonstrations of the negative volatility risk premium is that the fund has declined over 99% in the first seven years of its existence.

EXHIBIT 30.5 Quantifying the Volatility Risk Premium, June 1994–June 2012

Option Market	Average One-Month Implied Volatility	Average One-Month Realized Volatility	Implied Volatility—Realized Volatility Difference	Average Annual Excess Return to Selling Straddles	Standard Deviation of Short Straddle Returns
Equity Indices	20.3%	18.1%	2.2%	3.9%	3.9%
Commodity Futures	37.4%	33.0%	4.4%	6.1%	5.2%
Currencies	10.3%	9.4%	0.9%	1.2%	1.7%
10-Year Interest Rate Swaptions	23.4%	20.4%	2.9%	1.3%	1.0%

Source: Rennison and Pedersen (2012).

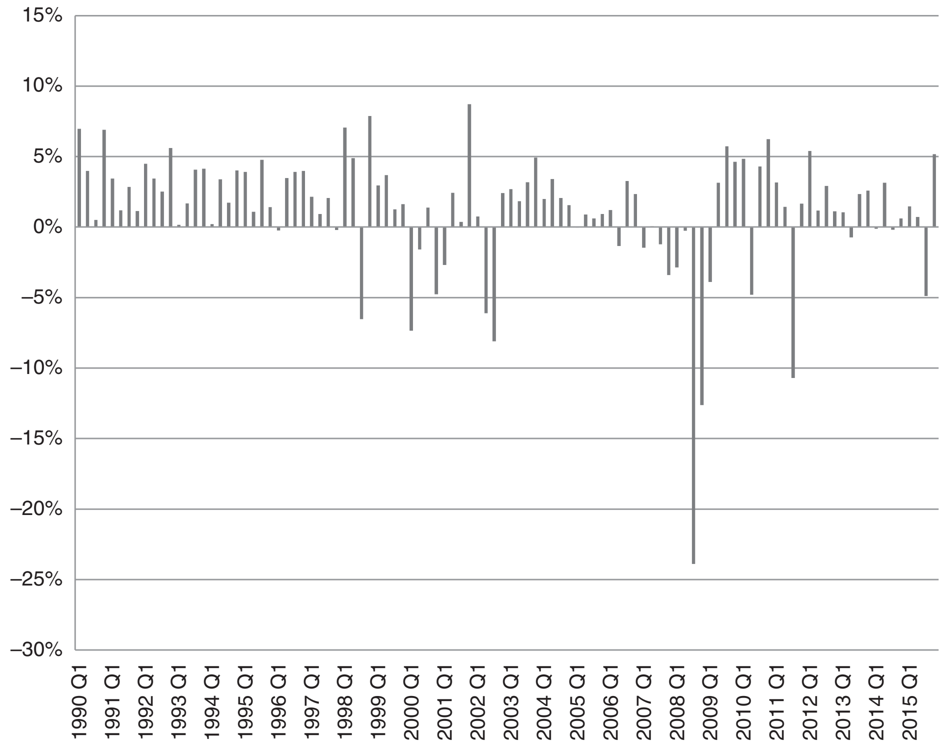


EXHIBIT 30.6 Quarterly Average 30-Day Richness of S&P 500 Options Implied Volatility Minus Realized Volatility, January 1990–December 2015
Source: Black and Szado (2016).

As illustrated by Black and Szado (2016) in Exhibit 30.6, the implied volatility of S&P 500 options exceeded realized volatility in approximately 90% of all calendar quarters since 1990. This strategy has posted gains since 1990, but experienced seven consecutive quarters of losses during the global financial crisis of 2007–9.

30.2.9 Dynamics of the Volatility Risk Premium

If a factor exposure provided consistent positive returns in all market conditions, it would lead to arbitrage. Thus, factor exposures should tend to generate gains in some markets (e.g., bull markets) while experiencing losses in others. Over several economic cycles, we should expect positive average returns to factor exposures that contain positive market beta (or other sources of systematic risk) and negative average returns (relative to the riskless rate) to factor exposures that contain negative betas. Risk premiums earned through traditional factor exposures tend to be positive during normal time periods, negative during periods of financial and economic distress, and positive over the long run. Returns from positive exposure to the volatility risk factor tend to be negative during normal economic periods, positive during periods of financial and economic distress, and negative over the long run.

There are large numbers of natural buyers who seek to hedge their risky investments by entering volatility derivatives with positions that are long volatility. If there is a limited number of traders willing to short volatility, then the risk premium must increase in order to attract traders willing to short volatility so that there will be a balance between the supply and demand for long volatility. It should be noted that shorting volatility exposes a trader to losses (i.e., tail risk) at the worst times: times of market distress. Therefore, there is reason to expect that the volatility risk premium will be substantially negative in equity markets as long as volatility itself is negatively correlated with equity returns.

30.2.10 Two Reasons Why Returns to Volatility Strategies Tend to Recover Quickly

The return to volatility strategies can be highly volatile. However, while long-only investments in stocks, credit, or commodities may take one to two years to recover from a drawdown, volatility strategies historically have recovered from a drawdown in less than a year. Two reasons are behind this shorter period to recover from drawdowns in volatility strategies. First, there is a high degree of mean reversion in realized volatility, which means periods of high volatility do not last very long, and, as realized volatility declines, suppliers of long volatility products recover their losses. Second, there is increased demand for long volatility products by investors in traditional asset classes after a crisis. The increased demand for long volatility is met through an increase in the anticipated volatility risk premium demanded by traders willing to take short volatility positions.

30.2.11 Two Reasons Why Volatility Mean Reversion Is Not Arbitragable

It is important to point out that while realized volatility may display a strong degree of mean reversion, this does not prove that the mean reversion can be used as the basis of highly profitable volatility trading strategies.

First, realized volatility is not directly traded. The related securities that are tradable are only indirectly related to realized volatility. For example, the settlement of the VIX futures contract is settled based on option prices that in turn depend directly on anticipated volatility rather than realized volatility.

Second, it is true that competition tends to prevent asset prices from displaying strong degrees of mean reversion in efficient markets since a tradable asset with a price that clearly exhibits mean reversion could be exploited to generate substantial and almost risk-free returns. However, nontradable market values such as interest rates, inflation rates, and volatilities can exhibit mean reversion since there is no ability for traders to directly exploit all patterns in rates. The CBOE Volatility Index (VIX) is a quantitative measure that displays a strong degree of mean reversion. But the VIX index itself is not tradable. However, the VIX futures contracts are tradable. As a result, the prices of the VIX futures contracts display far less mean reversion than the VIX itself.

30.3 MODELING VOLATILITY PROCESSES

The starting point for theory-based modeling of volatility derivatives is the specification of the underlying processes. Most valuation models for cash securities are based on continuous processes. Many volatility derivative models and some option models are based on jump processes. A **jump process** is often used to model a financial value that has potentially large discrete changes in the value rather having short-term changes that are all continuous and small. Models using jump processes are designed to better incorporate the large changes in market prices associated with major financial crises and liquidity events.

30.3.1 Volatility Processes with Jump Risk

Nossman and Wilhelmsson (2009) note that the negative expected return to long positions in futures on an index representing the implied volatility of options on the S&P 500 Index (VIX) is to compensate sellers for two volatility risk premiums: one for volatility diffusion and a second for volatility jump risk. To understand these two sources of risk, consider the following representation of returns on an equity index:

$$\frac{S_{t+\Delta} - S_t}{S_t} = \mu_t \Delta + \sigma_t \Delta \tilde{W} \tag{30.1}$$

Here, S_t is the equity price at time t, μ_t is the expected rate of return, σ_t is the standard deviation of the rate of return, $\Delta \tilde{W}$ is a random noise, and Δ is the length of time (e.g., one day). Equation 30.1 is a continuous-time diffusion process that does not allow for discontinuities or jumps in its values when large events occur. In some applications, such as the Black-Scholes option pricing model, the volatility parameter is assumed to be constant.

The valuation of volatility products must allow for volatility to change. The following is a general representation of changes in volatility:

$$\sigma_{t+\Delta} - \sigma_t = \gamma \Delta + \delta \Delta \tilde{Y} + \phi \Delta \tilde{J} \tag{30.2}$$

Here γ is the expected change in volatility, δ and ϕ determine the volatility of changes in volatility, and $\Delta \tilde{Y}$ and $\Delta \tilde{J}$ are random noises.

Note that Equation 30.2 permits one source of expected change in volatility ($\gamma\ \Delta$) and two sources of unexpected change. The values $\Delta \tilde{Y}$ and $\Delta \tilde{J}$ represent two different sources of uncertainty affecting volatility. $\Delta \tilde{Y}$ is a random noise that displays very small random positive or negative increments while $\Delta \tilde{J}$ is a random noise that displays large positive jumps. The risk arising from change in $\Delta \tilde{Y}$ is called volatility diffusion, while the risk arising from the jump component is called volatility jump. **Volatility diffusion** is the risk of volatility changes that represent the continuous accrual of small changes in the volatility of an asset through time. **Volatility jump** is the risk of potentially large periodic and sudden upward changes in the level of volatility.

30.3.2 Volatility Processes and Regime Changes

In financial markets, a **regime change** occurs when an observed behavior of a financial series experiences a dramatic shift. For example, a major macroeconomic shift such as a new central bank policy target combined with severe tightening of monetary

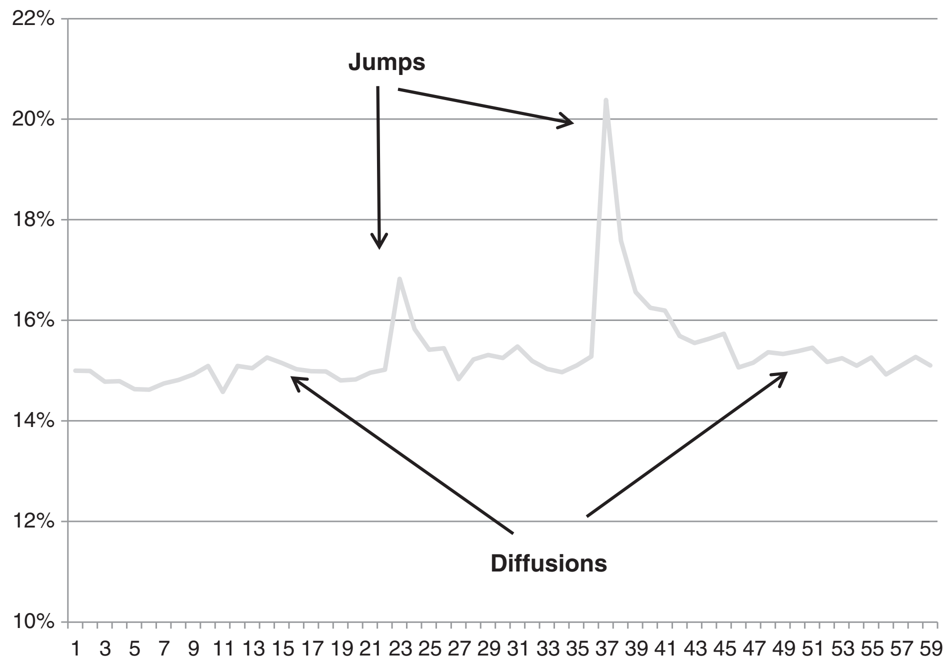


EXHIBIT 30.7 Example of a Volatility Process

policy could cause changes in the behavior of interest rates through time that might be described as a regime change.

In modeling traditional assets, it may often be reasonable to assume that the regime does not change. For example, the return-generating process of a large listed equity in a developed market might be little changed through time. However, volatilities and correlations in financial markets appear at times to experience substantial regime changes.

When volatility diffusion is the main source of uncertainty, volatility is lower and it follows a continuous path. However, when volatility jump is the source of risk, volatility will be higher and would follow a discontinuous path. Exhibit 30.7 displays an example of a volatility process driven by Equation 30.2.

Some investors model stock market volatility as a **mixture model or a regime switching model**, which models equity market volatility as some mixture of two return distributions. That is, there will be no jump component in Equation 30.2. However, there will be two different versions of Equation 30.2, with one having a much higher value for δ and one having a much lower value of δ. It is assumed that volatility switches between these two processes in a random fashion. For example, a low volatility market with a positive mean return may prevail for perhaps 80% of the time, and a high volatility market with a negative mean return may prevail much less frequently, the remaining 20% of the time. Because volatility clusters, the probability of staying within a current regime (either high or low volatility) is higher than the probability of switching to a new regime in the next time period. **Volatility clustering** occurs in a price series when large changes are likely to be followed by more large changes and periods of small changes are likely to be followed by more small changes.

30.4 VOLATILITY PRODUCTS

Equity market volatility, as well as volatility in other asset classes, reflects the enormous underlying risks inherent in modern economies. Markets are the institutions through which these economies gain the enormous benefits of trade. The ability to transfer and manage risk (i.e., trade risk) through financial markets is just as vital to the health of a modern society as the ability to trade commodities and products. Accordingly, modern economies need efficient and effective products to transfer risks such as volatility. A host of volatility products are now available to investors.

30.4.1 Four Common Features of Volatility and Fixed Income Products

As explained by Kuenzi (2007), volatility products and bonds have four common features:

1. Bonds and some volatility products are quoted in a similar manner, with quoted yields being used to calculate bond prices and quoted implied volatility being used to calculate options and volatility swap prices.
2. Bond prices are driven by interest rates, and volatility products are driven by volatilities, both of which tend to be mean reverting. Neither interest rates nor volatilities represent prices of traded assets, and therefore mean reversion can persist even in an efficient market. Rather, interest rates and volatilities drive the prices of assets. Because interest rates and volatility are mean reverting, they satisfy one of the main requirements of being a factor that generates a risk premium: They perform poorly during periods of financial stress and perform well during normal market conditions.
3. Both volatility and interest rates have a term structure, where long-term volatility and interest rates may be at very different levels than short-term volatility and interest rates. Shorter-term levels tend to be much more volatile, while longer-term levels have a lower volatility and are bounded to some extent by mean reversion toward the average level anticipated in the long run.
4. Sellers of volatility and buyers of bonds deserve to be compensated for the risk taken. Borrowers in the bond market are willing to pay a credit spread to compensate lenders for the default risk taken. Buyers of options are willing to pay a higher price for implied volatility than the expectation of realized volatility, either to hedge long positions in risky assets or to create short or levered positions in a simple way.

30.4.2 Variance Swaps Tied to Realized Volatility

To dispense with the daily rehedging of some investment strategies and to remove some of the path dependencies of vanilla options,[1] products exist where the returns are tied explicitly to some measure of realized volatility. Over-the-counter (OTC) variance swaps can be used to take a position on the realized variance of returns realized by the underlying asset. Variance futures are available on the EURO STOXX 50 Index, the Eurex, and the Chicago Mercantile Exchange (CME). The CME lists variance futures on a number of assets, including the S&P 500 Index, gold, EUR/USD, and natural gas, among others.

In a variance swap, one party receives the annualized variance realized by a specified asset, and pays a predetermined variance (a variance strike rate). For example, assume that in a variance swap the predetermined variance of the variance swap is 4.00 at the initiation of the agreement. Further assume that the realized variance that occurs in the underlying turns out to be 4.50. The buyer of the variance swap earns 0.5 times the notional amount of the swap, which is found by subtracting the predetermined variance from the realized variance. If the realized variance turned out to be 3.50, the variance swap buyer would owe the seller 0.5 times the notional amount of the swap. The predetermined variance strike rate in the variance swap is related to the level of implied volatilities from options. In other words, option implied volatilities provide a basis for determining an appropriate strike rate for a variance swap.

30.4.3 CBOE Volatility Index (VIX) Products

While variance swaps are commonly traded OTC and variance futures on a number of assets are listed, variance swap futures are relatively lightly traded in comparison to derivatives based on the CBOE Volatility Index. The **CBOE Volatility Index (VIX)** is a trademarked market-based approximation of the 30-day implied volatility of the S&P 500 that is calculated and disseminated in real time by the CBOE. The VIX serves as a direct underlier to futures contracts and to options. We provide detail on VIX-related products because trading based on the VIX represents a large portion of volatility trading and because they are similar to other products offered throughout the world. VIX-related products have achieved enormous levels of popularity in the wake of the financial crisis of 2008–9.

30.4.3.1 Computation of the CBOE Volatility Index (VIX) The VIX index is designed to represent the implied volatility of an option on the S&P 500 that has exactly 30 days to expiration. Since options on the S&P 500 have exactly 30 days to expiration only once per month and because only on rare occasions do options trade exactly at-the-money, it is necessary for the CBOE to use a weighted average of actual S&P 500 options. The exact formula relies on the formula for the valuation of a variance swap rate rather than an option pricing model. In order to arrive at the VIX index, the variance swap rate is calculated for two expiration dates: immediately prior to and after 30 days. The VIX index is then calculated as the square root (to convert a volatility-like value from a variance) of the interpolated values of the two variance rates.[2] All strikes for a maturity are used in the calculation for VIX, with the exception of any options with no bid price and any options with smaller deltas than those options with no bid price.

30.4.3.2 Futures Contracts on the CBOE Volatility Index (VIX) There are two major VIX derivatives based directly on the VIX: futures contracts and options. The futures contracts on the VIX allow market participants to take long or short positions in futures contracts with specific settlement dates. At the settlement date, each futures contract is settled based on the value of the VIX as computed on the settlement day based on a 30-day horizon. A long position in a single futures contract over the life of the contract will receive $1,000 for every point by which the VIX at the settlement date exceeds the price at which the contract was entered, or will pay $1,000 for every point by which the VIX falls short of the contract price. For example, 10 contracts entered on the short side at a price of 16.00 will require the trader to pay $5,000 (10 contracts × $1,000 per point × 0.50 points) on a mark-to-market basis if the contract settles at 16.50. The futures contract is available with monthly settlement dates up to roughly one year and weekly settlement dates up to six weeks.

Note that in all of the days prior to the settlement date of a futures contract the time horizon of the volatility exceeds the 30-day horizon on which the VIX is based. For example, consider a VIX futures contract with a settlement date of June 20. On June 20 the contract will settle based on option prices that have an expiration date 30 days later (July 20). Thus settlement will ultimately occur on 30-day implied volatility. However, prior to settlement, say on May 1, the futures contract with a settlement date of June 20 is based on options that expire in 80 days (i.e., on July 20). Thus the relevant volatility for the contract is the implied volatility between June 20 and July 20 even though the futures contract is trading on May 1.

30.4.3.3 The VIX Term Structure and Its Slope as a Proxy for Portfolio Insurance VIX futures contracts with different settlement dates have prices that reflect different windows of volatility. The **VIX term structure** is the relationship between the prices of VIX futures contracts and their settlement dates, usually expressed as a graph. The term structure of VIX futures contract prices can take on dramatic shapes based on volatility expectations and risk premiums involving a spectrum of prospective time intervals.

Usually, the VIX futures curve is upward sloping, or in contango: The front month VIX future is higher than the underlying VIX index, and the deferred VIX futures are higher still. Since the futures must converge to the value of the index, this characteristic of the futures curve can result in consistent losses to long positions in VIX futures, as shown in Exhibit 30.8, unless the VIX index increases. These

EXHIBIT 30.8 Return and Risk of VIX and S&P 500, 2006–15

Daily Observations, 2006–15	VIX Cash	VIX Short-Term Futures	VIX Mid-Term Futures	S&P 500
Annualized Return	4.20%	–40.60%	–12.30%	5.10%
Annual Standard Deviation	120.10%	64.70%	32.30%	20.80%
Skewness	1.31	0.85	0.56	–0.08
Kurtosis	6.15	3.56	3.29	10.56
VIX Beta	1.00	0.48	0.21	–0.13
S&P 500 Beta	–4.31	–2.35	–1.16	1.00

Source: Author's calculations, adapted from Black (2012).

losses are similar to insurance premiums that car owners or homeowners pay to insure against losses. In this case, long positions in VIX futures act as insurance for those investors who are using the futures contracts to hedge long positions in equities and other similarly risky assets. The insurance will pay off if there is a spike in the VIX, which is normally associated with a drop in risky asset prices. In other words, the slope of the VIX futures contracts represents the market price of insurance against spikes in the VIX and declines in prices of risky assets. Being long the VIX should be expected to be very profitable as a financial crisis or liquidity event unfolds.

30.4.3.4 The Determination of the S&P 500 VIX Short-Term Futures Index The performance of positions in VIX futures contracts varies substantially based on the settlement date of each contract. As with all futures contracts, hypothetical returns to a futures strategy depend on the settlement date selected and the procedure with which the contracts are rolled.

Investors in S&P 500 volatility tend to utilize a VIX benchmark known as the S&P 500 Short-Term VIX Futures Index to provide a measure of returns that uses a constant maturity. The **S&P 500 Short-Term VIX Futures Index** is a benchmark based on the estimated performance of a hypothetical VIX futures contract with a fixed time to settlement of 30 days formed using a time-weighted combination of the prices of the front and second month VIX futures contracts. The weights of the two contracts (one with less than 30 days and one with more than 30 days) are adjusted each day to linearly approximate the price of a 30-day futures contract. The weighting moves away from the shorter-term contract toward the longer-term contract as the times to settlement of both contracts diminish. The result roughly simulates a rollover strategy that rolls 1/30th of the portfolio into the next contract each day (ignoring holidays and weekends for simplicity) in order to keep the weighted average settlement date at a 30-day horizon. The linear interpolation of the two contract prices is depicted in Equation 30.3:

$$\text{30-Day Hypothetical Contract Price} = \frac{P_s \times T_s}{T_1 - T_s} + P_1 \frac{30 - T_s}{T_1 - T_s} \tag{30.3}$$

where P_s is the price of the shorter-term contract, T_s is the number of days to settlement of the shorter-term contract, P_1 is the price of the longer-term contract, and T_1 is the number of days to settlement of the longer-term contract.

For example, suppose that the nearby VIX futures contract has 10 days to settlement and a price of 20.10, while the first deferred contract has 40 days to settlement and a price of 20.40. A hypothetical contract with 30 days to settlement can be estimated using Equation 30.3 as the following linear combination of the contracts with 10 and 40 days to settlement, noting that there are 30 days between the settlements of the two contracts:

$$\text{30-Day Hypothetical Contract Price} = \frac{(20.10 \times 10)}{30} + 20.40\frac{(30 - 10)}{30} = 20.30$$

APPLICATION 30.4.3.4

Consider the information on two futures contracts over a two-day period in Exhibit 30.9.

EXHIBIT 30.9 Futures Contracts over a Two-Day Period

	Nearby Contract	Deferred Contract
Days to Settlement on Monday	15	45
Contract Price on Monday	15.00	16.20
Days to Settlement on Tuesday	14	44
Contract Price on Tuesday	14.10	15.00

Calculate the price of a hypothetical 30-day contract on Monday and Tuesday, and use the prices to calculate the price change between the 30-day hypothetical contracts on the two dates.

Using Equation 30.3, the hypothetical contract is priced at 15.60 on Monday (note that it is an equally weighted average since it lies equally between the two in time), whereas at 14.58 on Tuesday it lies closer to the deferred contract as the times to settlement of the contracts diminish and the weight on the longer contract increases. The price change of the hypothetical 30-day contract is –1.02.

There is also a Mid-Term VIX Futures Index that holds a combination of the fourth, fifth, sixth, and seventh month VIX futures contracts. Notice the large risk premium paid by investors taking long positions in the Short-Term VIX Futures Index contract, which can exceed 3% for each month, as shown in Exhibit 30.8. Long positions would earn a large return during a stock market crash such as occurred in 2008, but have a very high cost of carry. Positions in inverse, or short volatility, exchange-traded products have earned a positive return over some extended periods of time. Investors can access each futures contract separately in the futures markets, or trade exchange-traded products linked to the short-term or mid-term indices.

30.4.3.5 Options, Exchange-Traded Notes, and Other VIX-Related Products A spectrum of liquid options exist on the VIX. The options contracts settle based on the value of the actual VIX index on the option expiration date. Like the futures contracts on the VIX, the settlement (expiration) dates are weekly for up to six weeks and monthly up to roughly one year.

A spectrum of ETNs is available to provide exposures similar to long and short positions in VIX futures contracts. The ETNs differ by long versus short exposures, leveraged versus unleveraged exposures, and short-term or mid-term exposures. The VXX is a massive unlevered ETN that provides long exposure similar to the S&P 500 Short-Term VIX Futures Index. XIV is a large unlevered ETN that provides short exposure to the same index.

While the S&P 500 VIX asset complex offers the most diverse and liquid product universe, the methodology used to calculate the S&P 500 VIX index has been applied to a number of other assets globally. Moran (2014) explains that dozens of volatility-based indices are now available, with more than 40 exchange-traded products tracking several of these indices. The CBOE currently publishes volatility indices on equity markets (NASDAQ, S&P 500, Dow Jones Industrial Average, Emerging Markets, EURO STOXX, DAX, CAC, India, Hang Seng, single stocks); commodities (oil, corn, silver, gold); and currencies (Japanese yen and euro). There even is a "VIX of the VIX," the VVIX, which measures the volatility of VIX options.

30.4.4 Correlation Swaps

A **correlation swap** is a derivative that transfers the risk between the buyer to the seller that the actual average correlation among a specified set of individual stocks will differ from the swap's strike correlation.

30.4.4.1 Mechanics of a Correlation Swap A correlation swap has a fixed strike correlation paid by the swap buyer. The payment paid by the swap seller is based on the actual average realized correlation (with the average based on market weights) among the returns of the prespecified assets. For example, if the strike correlation rate of a correlation swap is 0.35, then the swap would require the swap buyer to pay 0.35 and receive the actual average realized correlation so that high realized correlations cause a net payment to the buyer from the seller and vice versa. The realized correlation is a market-weighted average of the return correlations between each pair of individual assets specified in the portfolio or index. The variable payment from the seller to the buyer is based on the average correlation rate multiplied by the notional value of the swap, whereas the fixed payment from the buyer to the seller is based on the strike rate times the notional amount.

APPLICATION 30.4.4.1

A correlation swap based on four assets has a strike correlation rate of 0.40 and a notional value of $100,000. Three of the assets experience actual realized return correlations between each other of 0.50. One of the assets experiences returns that have zero correlations with the other three assets. The market weights of the assets are all equal. Calculate the net payment to be made and identify whether the net payment is received by the buyer or the seller.

The four assets have six realized correlations. Three of the correlations are 0.50 and three are 0.00. Given that the asset weights are equal, the average realized correlation is 0.25. The payment amount is $(0.25 - 0.40) \times \$100{,}000 = -\$15{,}000$. The payment is made by the buyer to the seller.

Kelly (1994) defines the average correlation measure and shows how implied correlation tends to be overpriced versus historical correlation. He also shows how the average correlation relates to the implied volatility smile.

30.4.4.2 Modeling the Price of a Correlation Swap The price of a correlation swap is driven by the market-weighted average correlation between each pair of components, which is related to the implied volatilities of the index components and the index itself. Specifically, we know that the variance of the rate of return on a portfolio is given by the weighted average of variances as well as the weighted average of correlations:

$$\sigma^2_{\text{Index}} = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i}^{N-1} \sum_{j>i}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{30.4}$$

Here σ^2_{Index} is the variance of the rate of return on the index, w_i is the weight of security i in the index, and ρ_{ij} is the correlation between security i and j. If we replace pairwise correlations with the average correlation, ρ_{Average}, Equation 30.4 will be approximately correct; that is,

$$\sigma^2_{\text{Index}} \approx \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i}^{N-1} \sum_{j>i}^{N} w_i w_j \sigma_i \sigma_j \rho_{\text{Average}} \tag{30.5}$$

If we solve for average correlation, we have:

$$\rho_{\text{average}} \approx \frac{\sigma^2_{\text{Index}} - \sum_{i=1}^{N} w_i^2 \sigma_i^2}{2 \sum_{i}^{N-1} \sum_{j>i}^{N} w_i w_j \sigma_i \sigma_j} \tag{30.6}$$

30.4.4.3 Motivations to Correlation Trading Correlation trading centers on the relationship between the prices of options on portfolios (indices) and the prices of options on the individual assets in the portfolio or index. The relationship between the options pricing, in turn, depends on the return correlations between the individual assets because correlations drive the relationship between the total risk of a portfolio and the average total risk of the portfolio's constituent assets.

A similar strategy to a correlation swap is a strategy that is short index options and long options on the components of the index. This option trade will benefit if there is a decline in correlation. Correlation swaps allow investors to gain this exposure more directly. The buyer of a correlation swap agrees to pay a specific correlation and receive the market-weighted average correlation realized among the constituents of an index, for a specific dollar amount per correlation point.

As correlation typically rises during market turmoil, owning a correlation swap can be expected to be profitable during market downturns. Due to this dynamic, the price of a correlation swap is typically higher than the recent realized correlation in the market; selling correlation has often been a profitable strategy utilized by volatility funds.

30.4.4.4 Dispersion Trades A dispersion trade attempts to harvest the correlation premium. A dispersion trade is a long (or short) position in an index option and the opposite positions in the options of the index's constituent assets in which all of the option positions are hedged to be delta-neutral. The profitability of the trade will be based on the degree of dispersion between the returns of the index's components, which in turn is driven by the return correlations between the components.

Historical evidence is that the correlation implied by actual options prices is typically higher than the correlation actually realized by the underlying components. A trade to earn this spread, then, is to buy volatility on the index components (by buying options on each component) and sell volatility on the index itself (by selling the index option). For example, a dispersion trade on the NASDAQ-100 would require buying delta-neutral option positions on each stock in the NASDAQ-100 in an amount equal to the weight of each equity in the index, and then selling a delta-neutral option position on the NASDAQ-100 index.

30.4.5 The Common Theme in Volatility Trading

In terms of the pricing of volatility and volatility products, one common theme abounds. Consistent with asset pricing theory, investors are willing to pay for exposure to volatility, variance, and correlation, since all have a negative correlation to the broader equity market. As seen, passive long exposures to any of these instruments produce a negative return over time. However, similar to any other hedging product, their properties should be evaluated within a portfolio context to determine if they provide an adequate hedge and at what price. While this dynamic pervades the volatility landscape—in general, buyers of volatility lose money over time—volatility funds will include long volatility positions in their portfolios.

30.5 OPTION-BASED VOLATILITY STRATEGIES

Some option-based volatility strategies were discussed in the previous section in the context of similar non-option-based strategies. This section provides details on three major types of option-based strategies related to volatility trading.

30.5.1 Vertical Intra-Asset (Skew) Option Spreads

A **vertical spread** is a combination of long calls and short calls (or short puts and long puts) having the long options with one strike price (the long leg) and the short options with a different strike price but with the same expiration dates. The spread is intra-asset when it involves options from the same asset. The vertical spread is a type of skew spread.

30.5.1.1 Mechanics of a Vertical Spread While a vertical spread that is created using the same number of long and short options is a pure vertical spread, a **ratio spread** is a vertical spread with unequal numbers of long and short option positions. For example, a pure vertical spread would be long 10 March 100 strike calls and short 10 March 120 strike calls, while a ratio vertical spread may be long 10 March 100 strike calls and short 20 March 120 strike calls. In this section, we will study pure vertical spreads, and we will refer to them simply as vertical spreads.

At expiration, a vertical spread will have a minimum value of zero, when both options expire out-of-the-money. The maximum value will be equal to the difference between strike prices when both options are in-the-money. Depending on the options and the strike prices that are used, the position will have exposure to directional moves in the market. For example, buying an ATM put option and selling an OTM put option is a bear spread that will benefit if there is a decline in the value of the underlying asset. Also, because there is a negative relationship between implied volatility and changes in the value of the underlying asset, the position is long volatility. On the other hand, buying an ATM call and selling an OTM call will create a bull spread position that will benefit from a rise in the price of the underlying asset.

30.5.1.2 Vertical Spreads and Volatility Skews Volatility skews emanate from differences between implied volatilities across options that differ by strike prices. To take advantage of volatility skew, the traditional strategy of vertical spreads is to attempt to profit by selling OTM options that have a higher implied volatility than the ATM options. Using puts, the trader sells an OTM put and buys an ATM put (or using calls, the trader sells an OTM call and buys an ATM call). The skew is the difference between the OTM and ATM volatilities and is illustrated in Exhibit 30.3.

30.5.1.3 Vertical Spreads with Delta Hedging A common strategy is to use vertical spreads in equities while hedging delta risk. The strategy more commonly sells OTM puts and buys ATM puts. There is a tendency for volatility to expand as the price of the underlying drops, which by itself benefits the trader because the trader is long volatility.[3] To create a pure volatility play, the entire position can be delta hedged to remove the trader's exposure to changes in the underlying asset. In this case, the trader is attempting to benefit from two sources of potential returns. First, the trader will benefit if the implied volatility spread between OTM and ATM options narrows. Second, the position is delta hedged and net long volatility, and therefore will benefit if there is an increase in volatility. Noting that the ATM option that is being purchased is more expensive than the OTM option that is being sold, the primary cost of this position is the time decay in the value of the options. When realized volatility is higher than implied volatility, the large number of delta hedging trades will earn greater profits than the time decay of the option. Of course, when the differential continues to widen or the delta hedging underperforms the time decay, the positions will post losses.

30.5.2 Horizontal Intra-Asset (Skew) Spreads

The second type of intra-asset spread, a horizontal spread, differs in terms of the maturities or tenors of its options. A **horizontal spread** is a combination of long calls and short calls (or short puts and long puts) having the long options with one expiration date (the long leg) and the short options with a different expiration date but with the same strike prices. The spread is intra-asset when it involves options from the same asset. The horizontal spread is a type of skew spread.

While vertical spreads tend to have relatively significant exposures to directional moves in the underlying assets, horizontal spreads have much lower directional exposures. Horizontal spread returns are driven by changes in relative implied volatilities between options of different tenors, which in turn are often driven by the timing of new information such as an event.

An example of a trade using a horizontal spread occurs when a stock is set to release earnings in the next week, and the implied volatility of one-month options has risen substantially more than the implied volatility of three-month options. A trader may sell the one-month options and buy the three-month options in anticipation that the spread between the two implied volatilities will narrow after the earnings announcement due to a decline in the volatility of the short-term option. The trader will profit if the spread between the implied volatilities of two options narrows following the announcement if the price movement of the stock does not cause larger losses due to the position's negative gamma.

The previous trade example was based on a trader who anticipated that differences in the implied volatilities across option tenors would narrow. The trader sold short-term options and bought longer-term options to benefit from smaller differences between implied volatilities. Alternatively, a spread may be entered where one-month options are purchased and longer-dated options are sold. In this instance, the trader's strategy is to profit from an increase in the implied volatility of the short-term option relative to the long-term option.

If the trader is selling short-term puts and buying long-term puts, then the trader may be viewed as collecting an expected profit for providing insurance to those investors who wish to protect themselves against the impact of new information such as an earnings announcement. On the other hand, if the trader is selling short-term calls and buying long-term calls around an event date, the trader is providing a service to option buyers with a short time horizon who want to take a leveraged position in anticipation of a major move in the price of the underlying asset.

Unlike the vertical spread where one is able to calculate the maximum gain and loss from the position beforehand because the options share the same expiration dates, such estimates are less certain in the case of horizontal spreads because the expiration dates differ. Because of the differences in tenors, it is less clear what the payoff from the long-term option will be after the short-term option expires.

30.5.3 Inter-Asset Option Spreads

Traders often analyze the volatility spread between options on two different assets. An inter-asset option spread involves a long option position in one asset and a short option position in another asset. For example, the spread can be between two equities in the same sector, two currency pairs, or two commodities. The spread can be entered due to a catalyst in one leg of the trade—an earnings announcement, a central bank meeting, or an important geopolitical event—or the spread may be entered on the basis of analyzing past relationships between the two.

Trades are based on anticipated changes in the relationship between the implied volatilities of options on the two assets. Traders will establish long option positions on the asset with implied volatility that the trader expects will rise (relative to the implied volatility of the other asset) and will establish short option positions in the other asset.

An inter-asset spread can be illustrated using two equity indices. Exhibit 30.10 depicts the three-month implied volatility of both the S&P 500 Index and the

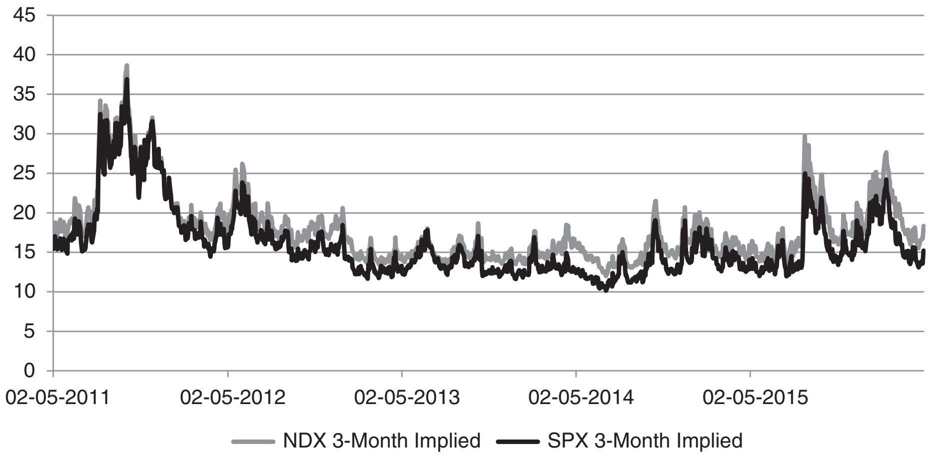


EXHIBIT 30.10 Implied Volatility of the S&P 500 and NASDAQ Indices

NASDAQ-100 Index from 2011 to 2015. Exhibit 30.11 shows the spread between the two. Note also that the spread reaches a high when the level of volatility is at its highs for the period, and hits a low when volatility is calm. Inter-asset option spreads would be established based on predictions of changes in the spreads between the implied volatilities of the two indices.

One reason for the observed relationship between the implied volatilities of the two indices is differences in their composition. The NASDAQ-100, as its name suggests, is composed of 100 stocks, while the S&P 500 Index contains 500. While both indices cover large-capitalization U.S. stocks, the sector composition is quite different, as the NASDAQ index has a 90% weight on stocks in the technology, consumer services, and health care sectors, compared to a weight of less than 60% on

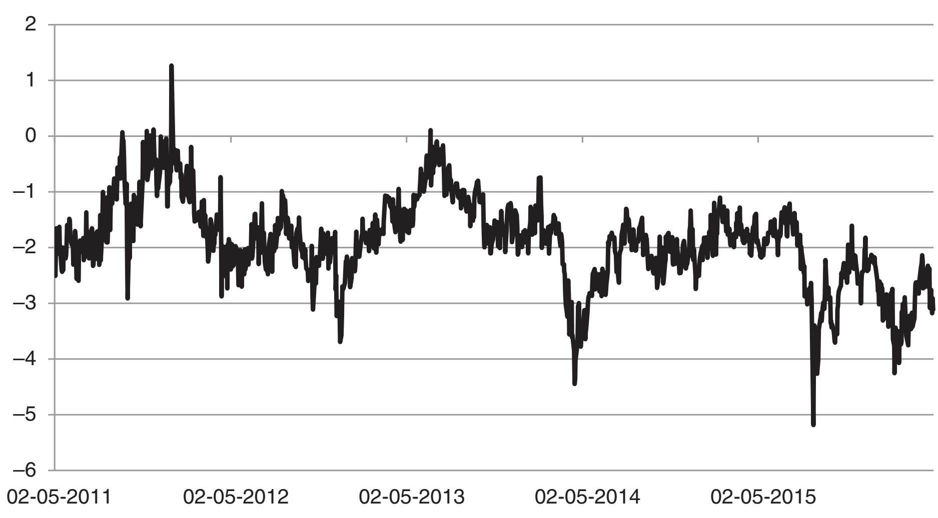


EXHIBIT 30.11 Volatility Spread between the S&P 500 and NASDAQ Indices

the same sectors in the S&P 500. Due to the composition of the two indices, the correlation of stocks within the NASDAQ index is typically higher than that of the S&P 500 Index.

Consider the case of this inter-asset spread in which the risk of changes in the value of each index has been hedged out to being delta-neutral using positions in the underlying assets (assuming volatility is constant). By being delta-neutral to each index, the strategy focuses on the relative implied volatilities between the two indices rather than the differences in their returns.

The implied volatility spread between the two indices essentially has two components to it. The first is the level of implied volatility of the stocks within each index. The second is the correlation between the stocks within each index. During periods of turmoil, the correlation among many assets typically rises. Lows in the spread between the implied volatilities of the two indices will tend to occur when the overall level of volatility is low, whereas highs will tend to occur when overall volatility is high. Accordingly, the side of the trade that is long the NASDAQ index option and short the S&P 500 Index option is long volatility and will gain in a period of market turmoil if the options are delta hedged.

30.6 VOLATILITY HEDGE FUNDS AND THEIR STRATEGIES

With increased investor attention to volatility products and the trading of these products, more focus has been placed on funds that primarily trade volatility.

30.6.1 Four Subcategories of Volatility Hedge Funds

Recently, Eurekahedge, in partnership with the CBOE, created fund indices that group volatility funds according to their portfolio composition and investment goals. Prior to this, all volatility funds typically were lumped together in a single volatility arbitrage category, regardless of their long or short volatility exposure (see the discussion of the HFRX Relative Value: Volatility Index in Chapter 19 of CAIA Level I). Given the negative correlation of short volatility funds with other strategies, the new indices provide a more granular look at the variety of volatility strategies.

The four CBOE Eurekahedge indices are Short Volatility, Long Volatility, Relative Value Volatility, and Tail Risk indices. These indices track funds with similar investment objectives and portfolio compositions. **Tail risk funds** are designed specifically to provide their investors with protection against large broad market declines. Long volatility strategies are a less aggressive variant of the tail risk strategy. While, as the name suggests, they should profit from increases in volatility, they may at times hold small or neutral volatility positions. Short volatility strategies chiefly are designed to extract the volatility or correlation risk premium inherent in stock indices and other assets. Relative value volatility strategies are designed to capitalize on volatility mispricings, and will buy volatility deemed undervalued and sell volatility deemed overvalued. While relative value volatility funds will hold long and short volatility positions, their overall position is often neutral.

30.6.2 Normalizing Vega Risk

Volatility funds hold positions in options and other securities that are subject to vega risk. Vega may be viewed as the sensitivity of an option price to a change in the volatility of the underlying asset or the implied volatility of the option. This section discusses normalization of vega risk. Vega can be defined as:

$$\text{Vega} = \frac{\partial \text{Call Price}}{\partial \sigma} \tag{30.7}$$

where σ is the volatility of the underlying.

Vega measures the response of an option price to a change of one unit in volatility, such as a change from 20% to 21%. Use of vega to control the risk of a portfolio of options implicitly assumes that all option implied volatilities change by the same absolute amount. Thus, if one option's implied volatility shifts from 10% to 11%, another option's implied volatility is implicitly assumed to shift by the same absolute amount, say from 20% to 21%.

A problem with using vega as usually defined is that the volatilities underlying different options frequently change by different magnitudes in predictable ways. For instance, an option with an initial implied volatility of 40% may be more likely to experience a larger absolute shift in volatility than an option with an implied volatility of 20%. The common approach to normalizing the vega for differences in the absolute size of volatility shifts across options is to adjust vega to represent identical relative changes. Thus the effects of the change in the implied volatility of one option from 10% to 10.1% would be modeled as being comparable to the change in the implied volatility of another option from 20% to 20.2% since the changes are equal on a relative basis.

Other normalization approaches can be used depending on the assumed relationship between changes in volatilities of options that differ by factors such as tenor or price.

30.6.3 Return Characteristics of Volatility Funds

Exhibit 30.12 lists the annual returns for the four volatility indices. Relative value funds are the most common type of volatility funds with 37 funds, more than the short (14), long (11), and tail risk (9) categories. The return history starts in 2005 for three of the four fund types, and begins in 2008 for the Tail Risk Index. The indices were created in 2015, and consist of the same group of funds for the entire period, which makes these returns subject to survivorship and instant history biases.

There was relatively low correlation across several pairs of these indices, which shows that there may be value to adding these indices together in a portfolio strategy. Also, note that the long volatility and tail risk indices had substantially greater variability of returns than the relative value and short volatility strategies. This relationship is evidenced in Exhibit 30.13, which contains the rolling 12-month correlations between each index and the S&P 500.

Caution should be used in interpreting the historical evidence in Exhibits 30.12 and 30.13 as being reliable forecasts of future performance, especially because the historical returns are observed over a relatively short period of time. Nevertheless,

EXHIBIT 30.12 Return Characteristics of Volatility Funds as Measured by Eurekahedge

	Relative Value Volatility	Short Volatility	Long Volatility	Tail Risk
2008	20.56%	–9.41%	45.81%	12.58%
2009	10.97%	19.10%	2.98%	–6.33%
2010	6.98%	11.35%	12.36%	0.10%
2011	5.46%	–1.20%	12.83%	16.39%
2012	8.81%	9.07%	0.27%	–21.21%
2013	6.04%	9.53%	–4.44%	–10.98%
2014	–0.36%	4.47%	1.58%	–3.22%
Annual Return	8.19%	5.77%	9.21%	–2.56%
Standard Deviation	6.42%	9.25%	16.93%	13.07%

Source: Eurekahedge, author's calculations.

the historical evidence indicates potentially attractive performance for relative value, short, and long volatility strategies in terms of both risk-adjusted performance and ability to diversify each other. It should be expected that tail risk volatility funds will perform poorly on an absolute basis over periods when markets do not experience substantial distress.

30.6.4 Relative Value Volatility Funds

As evidenced by the name, relative value volatility funds (Exhibit 30.14) will try to profit from mispricings and will take volatility positions in a variety of assets, often

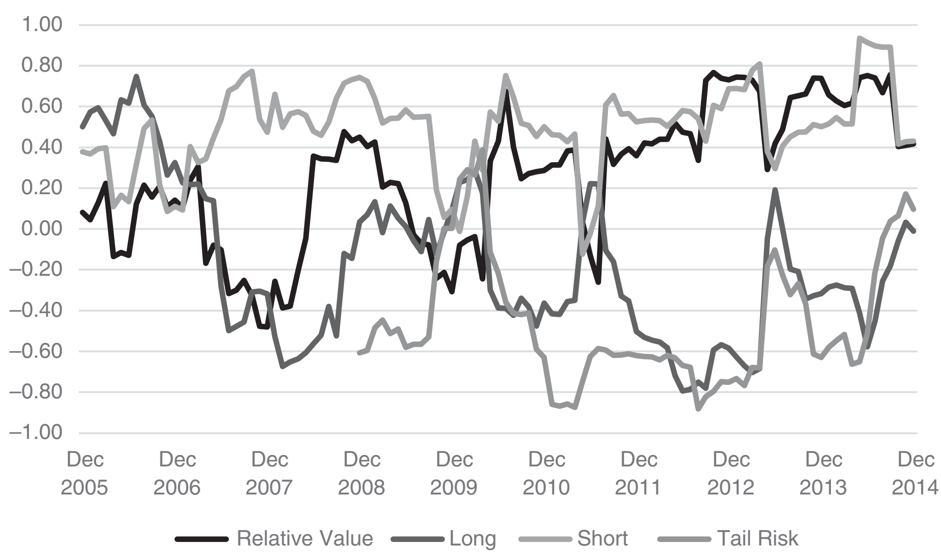


EXHIBIT 30.13 Volatility Fund Correlations versus S&P 500 Total Return
Source: Eurekahedge, author's calculations.

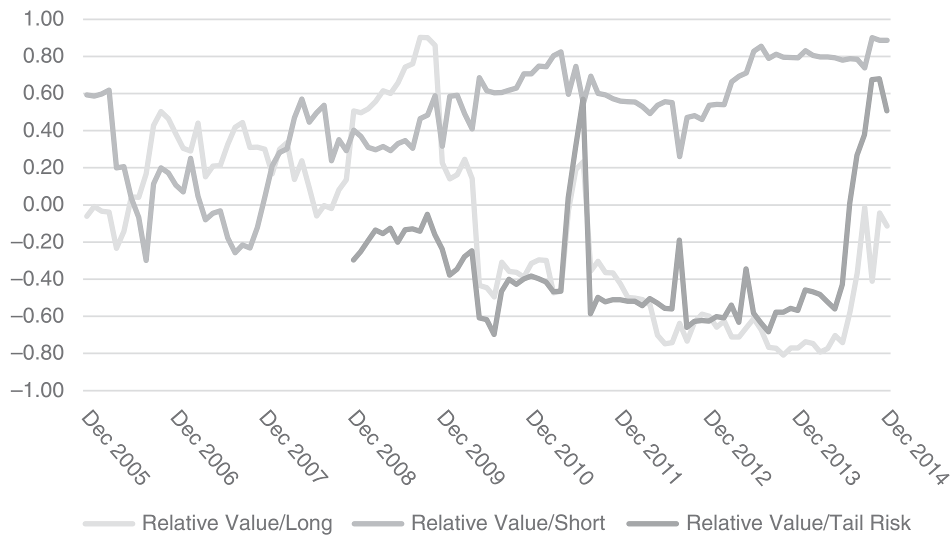


EXHIBIT 30.14 12-Month Rolling Correlations, Relative Value Volatility Funds
Source: Eurekahedge, author's calculations.

globally, with the goal of buying underpriced implied volatility and selling overpriced implied volatility. While the fund will hold a number of positions, the net vega exposure is small, as the fund will hold long and short positions. These funds may engage in a variety of strategies. They may buy or sell implied or realized volatility they perceive to be cheap or rich based on quantitative or fundamental measures, using options, variance swaps, and VIX products. They may trade intra- and inter-asset spread positions. They may also trade correlation, either through a dispersion trade or through a correlation swap. Many of the funds may also seek to capture the bid-ask spread in the options markets by making markets in a number of assets simultaneously and then managing the risks of the resulting positions.

30.6.5 Short Volatility Funds

Short volatility funds typically take positions designed to earn the volatility risk premium in stock indices, currencies, interest rates, or commodities. This premium is compensation for the risk that volatility increases, an occurrence that, again, occurs during equity market downturns or other events specific to the commodity or the currency markets. As an analogy, investors implementing a carry trade with long positions in high-yielding currencies will typically suffer losses during times of increasing systemic risk. Long volatility and tail risk funds will take long volatility or, less commonly, long correlation positions. These funds often research macroeconomic and microeconomic events to find stocks, sectors, indices, or other asset classes where future realized volatility may exceed the volatility implied by the option prices.

Short volatility funds will exhibit a bias toward carrying a net short implied or realized volatility position. While they may deploy both intra- and inter-asset spreads, these funds seek to profit chiefly from harvesting the volatility and correlation risk premium. As a result, positions held would include short variance swaps,

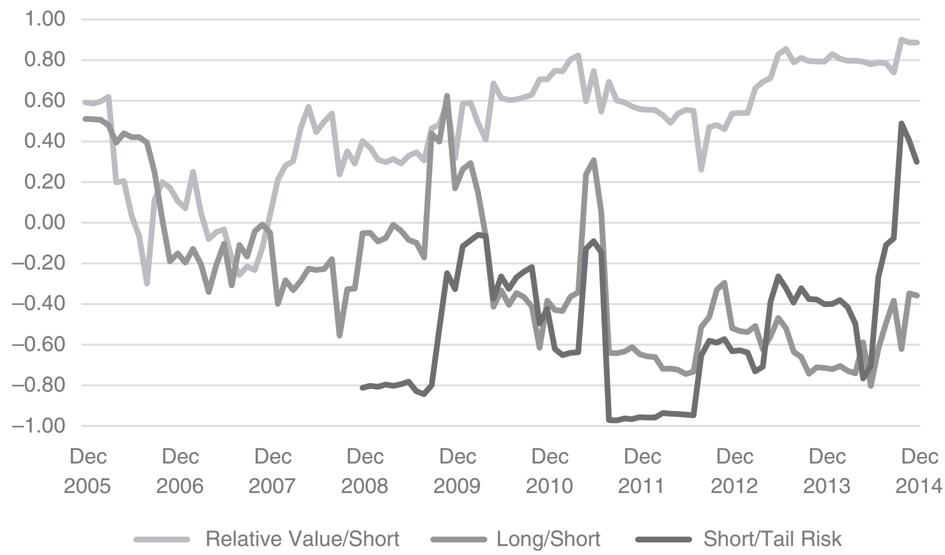


EXHIBIT 30.15 12-Month Rolling Correlations, Short Volatility Funds
Source: Eurekahedge, author's calculations.

short volatility products, dispersion trades, and correlation swaps. While the funds will tend to be short volatility, they can vary the exposure based on market conditions. Exhibit 30.15 contains the 12-month rolling correlations between short volatility and the other fund categories.

30.6.6 Long Volatility and Tail Risk Funds

Tail risk funds can be viewed as insurance products that protect their investors against broad market declines. While these funds gain during times of turmoil, investors should expect small declines in calm conditions as their long options or other protective positions decay. The profile of long volatility funds will be similar to that of tail risk funds in the sense that they should profit during market declines. However, they may not always require such a position, and they may hold short volatility positions that partially offset their long positions, so the gains during turmoil and the losses during calm markets may be smaller in magnitude. Long volatility funds will take positions similar to tail risk funds, but can vary their volatility exposure and hold more inter- and intra-asset spread positions, as these funds, unlike tail risk funds, do not have a mandate to provide protection at all times. As seen in Exhibit 30.16, the rolling 12-month correlations between long volatility and tail risk funds are high for much of the period.

Since these two categories of funds seek to protect against broad market declines, it is worth examining to what extent an allocation to either index benefits an equity portfolio. Exhibit 30.17 compares the S&P 500 Index to two portfolios: one with a 90% allocation to the S&P 500 Index and a 10% allocation to the Long Volatility Index, and one with a 90% allocation to the S&P 500 Index and a 10% allocation to the Tail Risk Index. The return volatility for both blended portfolios is lower, and

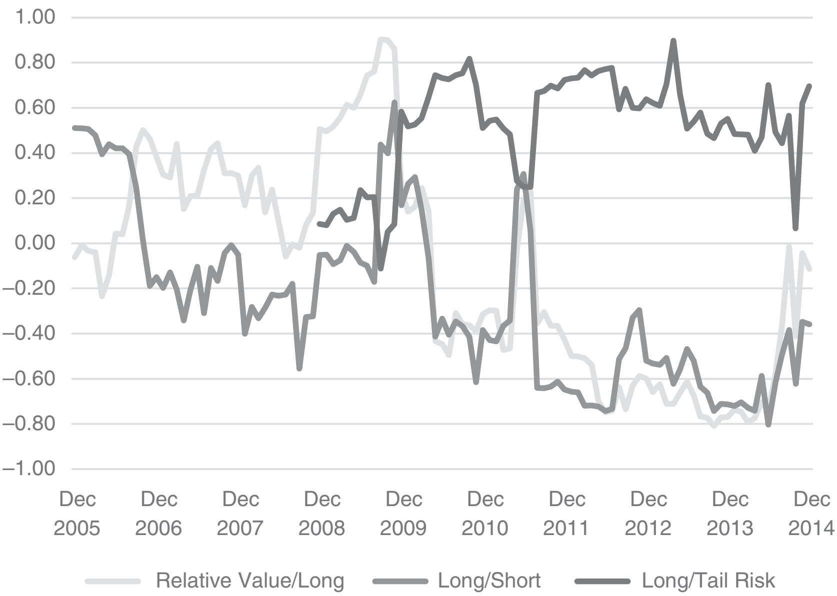


EXHIBIT 30.16 12-Month Rolling Correlations, Long Volatility Funds
Source: Eurekahedge, author's calculations.

the Sharpe ratio for both is higher. While the long volatility allocation appears to produce better returns, it is important to note two things. First, the first year of the Tail Risk Index, 2008, is disappointing, possibly due to the relative immaturity of these funds. Second, tail risk funds purportedly provide constant protection against black swans, while long volatility funds may not always be providing this protection. Nassim Taleb defines a **black swan** as an event or occurrence that deviates beyond what is normally expected of a situation and that would be extremely difficult to predict. Market events with large impact, such as the failure of Long-Term Capital Management or Lehman Brothers, tend to fit this definition of a black swan.

30.6.7 Tests of Four Volatility Strategies for Tail Risk Protection

Long volatility strategies can be effective hedges against tail risk events, but may come with a cost of substantial negative carry. Benson et al. (2013) examine four ways of implementing tail risk strategies. Their study focuses on finding strategies with positive returns during tail risk events, defined as those months when the equity market declined by 5% or more. Of the strategies that earned positive returns during these crisis months, the authors then seek to define the **performance drag**, which is the reduction in portfolio return when a specified strategy is added to a portfolio. The four strategies considered are:

1. Long volatility, including VIX futures and variance swaps
2. Low volatility equity, including HFRI short bias hedge funds and a portfolio of net short beta equities (long low-beta stocks and short high-beta stocks)
3. Trend following, as tracked by the Barclays CTA Index

EXHIBIT 30.17 Blending Tail Risk and Long Volatility Funds with an Equity Portfolio

	S&P 500	10% Long Volatility Index, 90% S&P 500	10% Tail Risk Index, 90% S&P 500
2008	–37.00%	–31.11%	–32.89%
2009	26.46%	24.17%	23.07%
2010	15.06%	15.00%	13.81%
2011	2.11%	3.29%	4.15%
2012	16.00%	14.42%	11.82%
2013	32.39%	28.24%	27.38%
2014	13.69%	12.45%	11.93%
Annual Return	7.27%	7.66%	6.57%
Standard Deviation	22.80%	19.65%	19.80%

Source: Eurekahedge, author's calculations.

4. Equity exposure management, which is a tactical long/short equity strategy based on moving averages or buying OTM equity index puts.

Strategy 1, which is described in this chapter, was found to have the greatest performance drag. While effective in hedging equity market declines, the long-term cost of long volatility strategies was simply too high. The criterion for a cost-effective tail risk strategy is a strategy that improves risk and return relative to moving to a cash position. If the performance drag is greater than that of a short-term high-quality sovereign debt investment, it should not be considered for addition to a portfolio of risky assets. The study concludes that the second through fourth strategies had the best performance over the period of the study, providing a hedge during times of equity market declines with the least performance drag. As seen in Exhibit 30.17 adding investments in the Tail Risk Index since 2008 has reduced the return and the risk of a long-only equity investment.

30.7 CONCLUSION

Since the financial crisis, investor attention to volatility products and funds has increased dramatically: The VIX index has spawned products around the globe that have become increasingly liquid, and volatility funds have differentiated themselves with regard to their objectives. Investors seeking to reduce portfolio risk may take long positions in volatility products, while those seeking to earn a risk premium could benefit from a diversified program of selling volatility. As new products emerge, volatility trading may offer potential for skill-based alpha.

NOTES

1. When using options to trade volatility, the path that the underlying asset takes can influence the risks to the investor. For example, an investor's exposure to volatility of the asset will decrease as the delta of an option tends to the extremes: When an option is very deep in- or out-of-the-money, any changes in volatility will have a very small impact on the option price.

2. Since 2014, the VIX calculation has included the use of SPX weekly options; prior to that, only the standard monthly expiries were used. The exact details can be found on the CBOE website at www.cboe.com.
3. In foreign exchange, this dynamic may exist in currency pairs commonly targeted due to the positive carry in the pair.

REFERENCES

Benson, R., R. Chapiro, D. Smith, and R. Thomas. 2013. "A Comparison of Tail Risk Protection Strategies in the US Market." *Alternative Investment Analyst Review* 1 (4): 32–47.

Black, K. 2012. "An Empirical Exploration of the CBOE Volatility Index (VIX) Futures Market as a Hedge for Equity Market and Hedge Fund Investors." *Research in Finance* 28.

Black, K., and E. Szado. 2016. "Performance Analysis of CBOE S&P 500 Options-Selling Indices." INGARM Working Paper.

Guiso, L., P. Sapienza, and L. Zingales. 2013. "Time Varying Risk Aversion." Working paper, University of Chicago Booth School of Business.

Kelly, Michael A. 1994. "Correlation: Stock Answer." *Risk* 7 (8): 40–43.

Kuenzi, D. 2007. "Shedding Light on Alternative Beta: A Volatility and Fixed Income Asset Class Comparison." In *Volatility as an Asset Class*, edited by I. Nelken, 61–82. London: Risk Books.

Lorenz, K., J. Lin, and J. Tsang. 2014. "The Enduring Case for High-Yield Bonds." TIAA-CREF Asset Management.

Moran, M. 2014. "Thirty Volatility Indexes: Worldwide Tools to Gauge Sentiment and Diversify Portfolios." *Journal of Index Investing* (Spring): 69–87.

Nossman, M., and A. Wilhelmsson. 2009. "Is the VIX Futures Market Able to Predict the VIX Index? A Test of the Expectation Hypothesis." *Journal of Alternative Investments* 12 (2): 54–67.

Rennison, G., and N. Pedersen. 2012. "The Volatility Risk Premium." *PIMCO Viewpoint*.

Sinclair, E. 2013. *Volatility Trading*. Hoboken, NJ: John Wiley & Sons.

Smith, D. A., and R. F. Whitelaw. 2009. "Time-Varying Risk Aversion and the Risk-Return Relation." Working paper, Stern School of Business, New York University.

CHAPTER 31

Hedge Fund Replication

This chapter discusses hedge fund replication methodologies and potential applications of products developed using these methodologies. The subject of hedge fund replication was first examined by academics in the early 2000s while attempting to develop performance benchmarks for hedge funds. Later, in 2007, following initiatives by major investment banks (such as Merrill Lynch, Goldman Sachs, Credit Suisse, and Morgan Stanley) and other firms to introduce investable hedge fund replication products, there was renewed interest in the subject among academics and practitioners.

This chapter examines, from both a theoretical and an empirical standpoint, the respective benefits and limits of the three approaches to hedge fund replication: factor-based replication, payoff-distribution replication, and bottom-up replication (also called algorithmic replication). The chapter also examines the potential benefits that replication products could offer investors, and provides a brief summary of some of the empirical evidence regarding the performance of these products.

31.1 AN OVERVIEW OF REPLICATION PRODUCTS

Fundamentally, **hedge fund replication products** (also called clones or trackers) are created to capture the traditional and alternative betas underlying the expected return and risk of a hedge fund benchmark. **Alternative betas** refer to exposures to risk, risk premiums, and sources of return that are not normally available through investments in traditional assets—or, if they are available, are commonly bundled with other risks. For instance, publicly traded equities have exposures to a number of factors, including volatility and commodity price risks. These two risks are generally considered alternative sources of risk, but in the case of common stocks, they are bundled with pure equity risk, which dominates the behavior of common stocks. Other examples of alternative betas are risks associated with currency investments, momentum or trend-following strategies, and structured products, such as convertible bonds and certain tranches of asset-backed securities.

Available replication products are based on statistical techniques (e.g., factor-based and payoff-distribution approaches) or bottom-up (algorithmic) trading models. The latter approach attempts to trade the underlying securities in a manner consistent with the trading approach taken by most active managers within a particular strategy. To the degree that tracker products are designed to track benchmarks based on active managers, they are meant to capture a significant portion of the common strategy return of fund strategies and to reflect fund managers' common exposures to various traditional and alternative betas. Additionally, these tracker products are designed to capture the common strategy alpha earned by managers, as represented by the excess return of the fund strategy relative to a passive benchmark consisting of traditional sources of risk (e.g., MSCI World Index). In other words, if the overall strategy, rather than just the top managers, generates alpha, then replication products may be able to capture all or some of that alpha.

31.2 POTENTIAL BENEFITS OF REPLICATION PRODUCTS

Before discussing the various approaches to hedge fund replication, the potential benefits that one may receive from using a replication product must be examined. The term *using* is employed, rather than *investing*, because investors may benefit from these products even if they do not make any allocations to them, as replication products can lead to a better understanding of the underlying risks of hedge funds and allow investors to build better benchmarks against which to measure their hedge fund investments. The reason that investors may allocate to hedge funds in the first place needs to be understood. In other words, why is it of some potential value to replicate hedge funds?

As discussed in the CAIA Level I and Level II curriculum, to understand the potential benefits of investing in hedge funds, one must distinguish among various strategies. Some strategies have significant systematic risks and therefore allow investors to earn relatively high risk-adjusted returns through exposures to directional market risks. Equity long/short funds and some global macro funds fall into this category. The primary benefit that investors seek in investing in these return enhancers is to improve the risk-adjusted returns of their portfolios. This benefit can come from earning alpha or by investing in alternative beta exposures that are underweighted in an investor's traditional portfolio. The alpha is typically measured relative to the performance of all underlying beta exposures, whether they are traditional betas or alternative betas. Alternative beta exposures may include those taken by a global macro fund that invests in global equity and foreign currency markets while managing the risks of those investments. Liquidity risk is another important alternative source of return not available through investments in traditional assets. For example, liquidity risk is one of the primary sources of return to a distressed securities strategy. Finally, a time-varying traditional source of beta could be considered an alternative source of beta. For instance, a passive equity index has a constant and positive equity beta, while an equity long/short manager is likely to have a time-varying beta, which could even become negative during bear markets. Equity beta is considered a traditional source of return, while a dynamic beta that results from an actively managed portfolio is considered an alternative beta and, assuming that the manager has some market-timing skill, serves as a source of alpha.

Some strategies, however, have less systematic exposure to equity, fixed-income, or commodity markets, and are classified as risk diversifiers. Some commodity trading advisers (CTAs) and global macro funds, as well as most relative value funds, such as convertible and merger arbitrage, fall into this category. These funds provide returns that are not highly correlated with returns on traditional assets and can therefore help reduce the overall risk of a portfolio. Similar to return enhancers, a secondary benefit provided by risk diversifiers is exposure to alternative sources of risk and return. For instance, convertible bond arbitrage funds provide exposures to risks associated with implied volatility, credit, and illiquidity, earning a relatively low rate of return while reducing the overall risk of a traditional portfolio.

31.3 THE CASE FOR HEDGE FUND REPLICATION

Prior to the bursting of the Internet bubble and the resulting bear market that began in 2000, there was a widely held market belief that the returns to actively managed alternative investment funds were composed primarily of alpha (excess return above that available in an equally risky passive investment product) and small amounts of beta (return due to a strategy's exposure to investable market factors). Since then, the return streams associated with passive investable indices, active managers, and other alternative or structured products have severely challenged this belief system. In the traditional mutual fund arena, research continues to question the existence of manager alpha (e.g., see Bodie, Kane, and Marcus 2010). Even in the alternative investment area, where greater amounts of informational inefficiencies may be expected, studies have shown that the alpha of many hedge fund strategies is consistently positive but that there has been a general decline in overall alpha in the past decade.[1]

31.3.1 Estimating the Risk and Return of Funds of Funds

Exhibit 31.1 displays a one-factor estimate of the rolling beta and rolling alpha of the Center for International Securities and Derivatives Markets (CISDM) Fund of Funds Index, for which the S&P 500 Index is used as the benchmark. Both the beta and the

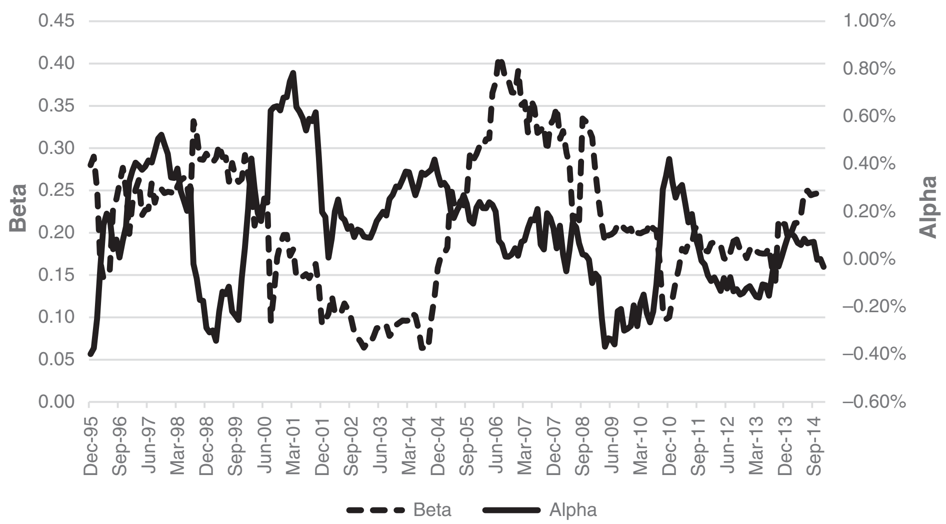


EXHIBIT 31.1 Twenty-Four-Month Rolling Estimates of Betas and Alphas of the CISDM Fund of Funds Index Relative to the S&P 500 Index, 1994–2014
Source: CISDM, Bloomberg, and authors' calculations.

alpha are estimated using a 24-month rolling window. The alpha is estimated using the following expression:

$$\overline{R}_{t,CISDM} - r_f = \alpha_t + \beta_t(\overline{R}_{t,SP} - r_f) \qquad (31.1)$$

Here, $\overline{R}_{t,CISDM}$ is the estimated 24-month mean return of the CISDM Fund of Funds Index, $\overline{R}_{t,SP}$ is the estimated 24-month mean return on the S&P 500 Index, β_t is the estimated beta of the CISDM index to the S&P 500, α_t is the estimated alpha of the CISDM index, and r_f is the monthly rate for 30-day London Interbank Offered Rate (LIBOR). All variables are estimated using a 24-month rolling window.

From Exhibit 31.1, it can be seen that since 1995, there has been significant variation in the beta of the hedge fund index. A rise is observed from about 2004 to 2007, and then a general decline since then. However, a general decline in the value of alpha is observed since 2001, with the most recent estimates hovering around zero.

31.3.2 Three Theories for Increased Beta and Decreased Alpha in Hedge Fund Returns

There have been various explanations for increases in beta and declines in alpha: changes in trading strategies, increased liquidity, technological advances that have all but eliminated information premiums, or regulatory intervention. These explanations have given rise to three alternative hypotheses:

1. The fund bubble hypothesis
2. The capacity constraint hypothesis
3. The increased allocation to active funds hypothesis

The **fund bubble hypothesis** assumes that successful hedge fund managers can earn substantially greater incomes than successful fund managers in the traditional space. Investment bubbles provide an opportunity for less skilled traditional fund managers to become hedge fund managers. The fund bubble hypothesis simply states that as the supply of investment capital to hedge fund space increases, so does the number of less qualified managers who enter the industry and provide inferior returns that dilute the overall performance of the industry.

The **capacity constraint hypothesis** argues that most alpha is a zero-sum game. Therefore, only a few managers can offer alpha on a consistent basis. In short, the growth in assets under management sharply reduces the per capita amount of alpha available in the marketplace. In addition, unless new strategies or sources of alpha are discovered, further declines in alpha can be expected.

The **increased allocation to active funds hypothesis** argues that as hedge fund investment becomes more popular, the performance of hedge funds will be adversely affected by the trading decisions of investors who have allocations to these funds as well as to traditional assets. In other words, the systematic risks or betas of hedge funds increase as more capital flows into them. In particular, during periods of financial stress, investors may be forced to liquidate both their traditional and their alternative investments, increasing the correlation between traditional and alternative asset classes.

The common theme flowing through these three hypotheses is that, although star managers exist whose ability to provide meaningful returns is not questioned, the investment management universe also includes managers who either are mandated to invest in accordance with a given benchmark or do not have the skill to surpass that benchmark. Note that the lack of manager outperformance is not necessarily manager induced. For example, some managers may be constrained by the mandates of their contracts; in this case, often the clients have made an asset allocation decision that requires the manager to remain within certain limits. Client intervention, however, plays a very limited role within absolute return vehicles, whereas the decline of alpha is pronounced and without many mitigating factors. The real questions in the replication or tracking literature remain the degree to which the representative benchmark is investable, and the degree to which the investable replication or tracker fund adequately reflects the risks and returns of the representative benchmark.

31.3.3 The Aggregate Alpha of the Hedge Fund Industry

Since hedge fund replication products attempt to capture the overall performance of hedge funds or hedge fund strategies, it is important to examine whether the hedge fund industry can generate alpha on an aggregate level. While there is ample evidence that some hedge fund managers are able to provide alpha, it is not clear whether the entire strategy can and should be able to generate positive alpha on a consistent basis. The main argument against the ability of the entire industry to generate alpha is rooted in the concept of market efficiency, which was discussed in detail in the CAIA Level I curriculum. However, if security markets are perfectly efficient, then paying positive fees to active managers implies inefficient markets for asset management. In other words, one cannot simultaneously assume that financial markets are dominated by rational investors who arbitrage away pricing inefficiencies and that there are irrational people who invest with professional money managers who, according to the efficient market hypothesis, do not add any value. Why should professional money managers exist at all, and why should some investors be willing to pay them substantial fees to manage their assets? The presence of active money managers seems to imply that markets are not efficient, as thousands of hedge funds, private equity funds, and active mutual funds earn substantial fees when, according to the efficient market hypothesis, they should underperform active strategies on an after-fee basis. However, if markets were highly inefficient, many more traders would enter the active money management business to earn a portion of those fees and, in the process, help make markets more efficient. Therefore, it seems that there must be a fine balance between efficiency and inefficiency. This leads to a situation called being efficiently inefficient.

Being efficiently inefficient presents the idea that markets are, on average, just inefficient enough to compensate managers and investors for their costs and risks, but not so inefficient as to present a large number of money managers with low-hanging fruit. Therefore, the flow of capital to actively managed funds is limited in a world that is efficiently inefficient. In such a world, competition among active money managers results in markets that are almost efficient but in which some inefficiencies exist that reward those who can identify and exploit them. In a market characterized as being efficiently inefficient, there exist a limited number of market inefficiencies that can be exploited by some money managers. However, finding the right manager takes time and resources, and therefore investors have to decide whether to spend search costs to find an active asset manager or allocate their capital to a passive strategy. At the margin, investors become indifferent between choosing passive investment products and searching for an active asset manager. If search costs are low such that investors can easily identify good managers, then more money is allocated to active management, and some pricing inefficiencies are arbitraged away.

Of course, there are investors who lack the resources, the patience, or the skills to search for good active managers, and as a result, they allocate randomly to both good and bad managers. In fact, one can argue that these asset allocators are more likely to invest with bad managers, as the skilled managers tend to have capacity constraints. The performance of these allocators will depend on their relative allocations to good and bad managers, but their overall performance after fees is likely to be worse than that of passive investors. In addition, if these allocators represent a relatively large portion of investors, then the overall performance of active managers is likely to be worse than that of passive managers. Therefore, whether replication products add value or not depends on the relative number of active allocators who are able to spend resources to eliminate unskilled managers from the pool of available hedge fund managers. However, if due diligence is too costly and finding skilled managers too uncertain, then unskilled managers will not be eliminated from the available pool, and the overall performance of the industry will suffer.

31.3.4 Replication Products as a Source of Alpha

It is commonly argued that replication products cannot possess alpha, because they are not managed by skilled managers. However, replication products that track a benchmark consisting of one or more actively managed funds may be able to capture the alpha provided by the benchmark. In the extreme case, if a replication product can perfectly replicate the return properties of the benchmark, then by definition it will capture the alpha provided by the benchmark as well. If the benchmark consists of a group of top managers, then the replication product may even be able to capture the alpha provided by top-tier managers. Whether in practice one can create such a perfect replication product will be discussed later.

As previously noted, the presence of alpha in aggregate hedge fund indices is directly related to the ability of investors who directly allocate to hedge funds to identify skilled managers and help eliminate unskilled managers from the available pool of hedge funds. Further, since hedge fund replication products carry lower fees than actively managed portfolios, the possibility of providing alpha to investors who allocate to these products increases. On the other hand, if the replication product cannot gain access to certain sources of returns used by managers in the underlying benchmark, then replication products will not be able to generate the level of alpha provided by managers. In short, whether replication products can generate alpha is an empirical question.

31.3.5 Replication Products as a Source of Alternative Beta

Replication products have the potential to give investors access to some of the alternative sources of beta. Whether this potential can be realized in practice depends greatly on the set of securities used to create the replication product. For example, if liquid publicly traded securities are used to create the replicating portfolio, then clearly the product cannot provide exposure to illiquidity risk. This in turn means that any return that can be attributed to illiquidity risk will not be present in the replicating product's performance.

As new passive investable products are created, replication products are given a greater chance to offer investors access to many sources of alternative betas. For example, exchange-traded funds (ETFs) based on convertible bonds and volatility have recently become available. These ETFs make it possible to create investable products that would track convertible bond arbitrage and volatility trading strategies. For instance, a strategy consisting of a long position in the convertible bond ETF and a short position in an equity index ETF could generate returns that are similar to a simple convertible arbitrage strategy. An interesting question that arises is whether these can still be considered sources of alternative beta now that they are available to all investors through traditional investment vehicles. To the degree that these new investment products bundle an alternative source of beta with other traditional sources of risk, they may still be considered alternative assets. For example, a convertible bond ETF provides exposure to the options embedded in convertible bonds while exposing investors to credit, equity, and interest rate risks. Some of these risks are commonly hedged away by active managers, and thus a convertible bond arbitrage fund may have little exposure to equity or interest rate risks. Therefore, the embedded option of a convertible bond and its implied volatility can still be considered potential sources of alternative beta. Finally, as discussed in the previous section, time-varying betas associated with traditional sources of risk could be considered alternative betas. Therefore, to the degree that a replicating product can capture the dynamic beta of a hedge fund, the product could be a source of alternative beta.

31.4 UNIQUE BENEFITS OF REPLICATION PRODUCTS

In the previous sections, it was argued that replication products have the potential to provide many of the benefits that motivate investors to allocate to hedge funds. Namely, they could act as both return enhancers and risk diversifiers. This section examines the benefits that are somewhat unique to replication products and are generally not offered by most hedge funds. These represent the strongest rationales for the development of various replication products and the reasons why some investors consider allocating to them.

31.4.1 Two Key Issues Regarding Fund Replication Benefits

Whether the unique benefits provided by hedge fund replication products make a strong case for increased allocations to these products depends primarily on the investor's initial reason for investing in hedge funds. If access to some unique sources of risk premium (e.g., illiquidity) and skills displayed by top-tier managers are the main reasons, then replication products will not be considered viable alternatives to hedge funds. On the other hand, if the goal is to capture the alpha and the beta that are represented by the benchmark that underlies the replication product, then these products may be attractive to some investors. The two key questions in this discussion are:

1. Can one identify top-tier managers a priori, and do managers display significant persistence in their performance?
2. Can hedge fund replication products track performance of various strategies during different market cycles?

Regarding the first question, academic and industry research provides mixed results. Some studies show that top-tier funds do display return persistence, but their outperformance is partially eroded through time. This is because capital flows to top-performing funds, leading to some erosion of their superior performance.[2] Other studies show no performance persistence among hedge fund managers—or, if there is any, it tends to disappear after a few months.[3] The second question is discussed later in the chapter.

31.4.2 Eight Potential Unique Benefits from Hedge Fund Replication

This section examines eight unique benefits that hedge fund replication products can potentially provide.

31.4.2.1 Liquidity as a Replication Benefit Tracker products invest primarily in liquid instruments (e.g., ETFs or futures) and are therefore able to offer liquidity terms (both purchase and redemption) to investors that most fund managers cannot match. Most replication products do not have lockup periods or the ability to erect gates to slow investor withdrawals. Though it is difficult to quantify the redemption premium required by fund investors, evidence from the secondary market for hedge fund investments shows that during periods of market distress (e.g., 2007–8), investors were willing to accept a discount of 20% of the net asset value (NAV) of their hedge fund investments in order to exit a fund.[4] Even though this reflects the cost of a liquidity premium during a market crisis, the cost of redemption liquidity—even under normal market conditions—is likely to be a few percentage points.

Hedge fund investors are bound by the redemption policies of the hedge funds. An investor in a hedge fund often has to manage liquidity by attempting to anticipate the behavior of other investors, who may decide to redeem quickly, leaving the long-term investor with the least liquid (and least attractive) remnants of a fund's portfolio. An investor who executes a replication strategy does not face these problems. The investor will enjoy whatever liquidity is on offer from the actual market and is not hurt if other investors decide to exit early.[5]

Hedge fund investors can achieve some degree of liquidity and reduce exposure to the behavior of fellow investors by using a managed account platform. The key advantage to a managed account is complete control by the investor. By pulling trading privileges, the investor has the ability to manage the cash and liquidate the account at any time. In theory, this gives the investor better than daily liquidity, as the account can be liquidated whenever the market is open. Of course, these advantages come at a cost. For example, one cost is the reduced size of the pool from which managers are selected. Many large managers do not accept managed accounts, and those that do require rather large minimum account sizes and have other additional administrative requirements.

31.4.2.2 Transparency as a Replication Benefit Replication products can afford to be highly transparent in terms of their trading strategies, including security holdings. Moreover, since replication products operate primarily in the most liquid segments of the markets (e.g., exchange-traded securities), trades implemented tend to have negligible price impact when trading is focused on the most liquid ETFs. Further, since trades are often primarily algorithmic, suppliers of tracker products can disclose a significant amount of information about the trading process. As discussed previously, hedge fund investors could receive a high degree of transparency by using a managed account platform. However, while such an approach increases transparency in terms of portfolio positions, it may not provide complete transparency in terms of the trading process of the portfolio manager.

31.4.2.3 Flexibility as a Replication Benefit Most replication products can be very flexible in terms of the risk profile they offer investors. An investor may be able to specify a particular hedge fund strategy benchmark that the product is designed to track, as well as its volatility and other desired statistical properties. These products are especially suited for use in separate accounts designed to meet individual investors' unique requirements. Funds of funds (FoFs) may be able to put investors' assets to work rather quickly by investing in a replication product.

31.4.2.4 Lower Fees as a Replication Benefit Replication products charge lower total fees than do hedge funds. The lower fees for replication products may offset the higher gross returns that individual hedge fund managers may generate, reducing the differential net-of-fee returns generated by both products. Further, once other costs such as market impact, due diligence, monitoring, and liquidity are taken into account, replication products can offer net returns that are comparable to those offered by individual funds and funds of funds. The results in Exhibit 31.2 for a given hypothetical level of return indicate that the higher expenses in both direct hedge

EXHIBIT 31.2 Gross and Net Return Comparisons

	Hedge Funds	Funds of Funds	Replication Products
Net return assumed to be earned by investors	7.00%	7.00%	7.00%
Management fee—fund level	2.00%	2.00%	1.00%
Performance fee—fund level	20.00%	20.00%	
Estimated operating expenses—fund level	0.40%	0.40%	0.40%
Management fee—FoF level		1.00%	
Performance fee—FoF level		10.00%	
Estimated operating expenses—FoF level		0.40%	
Gross return required to achieve comparable net return	11.15%	13.87%	8.40%
% of gross return earned by investors	62.78%	50.47%	83.33%

Source: Crowder, Kazemi, and Schneeweis (2011).

fund and FoF investments require a hedge fund–based product to produce a substantially higher gross return compared to that of a replication product, if investors are to receive the same net return. The higher expenses generally involved in hedge funds and funds of funds set a high gross return hurdle if their net returns are to exceed those of comparable replication products. In Exhibit 31.2, only 62.78% and 50.47% of the gross returns earned by hedge funds and funds of funds, respectively, flow to the investors. In contrast, 83.33% of the gross returns of replication products are earned by investors. Further, if these products were used in structured products, structuring fees applied to tracker investments would be significantly lower due to their transparency and liquidity.

31.4.2.5 Hedging as a Replication Benefit If replication products are created using liquid financial instruments, then it should be possible to short the replication product. This creates a number of potential opportunities for investors. For example, if an investor cannot reduce her allocation to a hedge fund, she might be able to hedge some of its risks by shorting a replication product that is designed to mimic its strategy.

31.4.2.6 Lower Due Diligence and Monitoring Risks as a Replication Benefit Recent experience has shown that hedge funds and funds of funds have significant due diligence and operational risk exposure. Replication products entail lower operational risk, imposing substantially lower due diligence and monitoring costs on their investors. In addition, hedge funds and funds of funds may deviate from their expected investment styles, exposing investors to style drift risk. Since replication products are calibrated to a defined hedge fund strategy benchmark, investors have minimal exposure to style drift.

31.4.2.7 Diversification as a Replication Benefit In some replication approaches, diversified factors, such as the returns on ETFs or futures contracts, are used to create portfolios that mimic the performance of hedge funds. This creates replication products that are highly diversified. Even when the product's positions are concentrated in a few economic sectors or countries, diversified instruments such as ETFs and futures are normally employed to gain these exposures. Diversification benefits offered by replication products also reduce the risk of both large drawdowns and high volatility associated with investing with a small number of managers or securities. Investors can achieve the same degree of diversification by investing in a portfolio of hedge funds or in a diversified fund of funds. However, both alternatives entail added costs in terms of either increased monitoring costs or extra layers of fees.

31.4.2.8 Benchmarking as a Replication Benefit The preceding benefits would accrue only to investors who allocate to replication products. However, the benchmarking benefit does not require any allocation to the replication product. If a replication product were able to capture the properties of the returns to a given hedge fund strategy, then it would represent an investable benchmark, which can be used by investors to estimate the value added by their managers. In addition, the investor may be able to negotiate incentive structures that are tied to the manager's performance relative to a replication product.

31.5 FACTOR-BASED APPROACH TO REPLICATION

As discussed in the chapter opener, there are three broad approaches to hedge fund replication: (1) factor-based, (2) payoff-distribution, and (3) bottom-up or algorithmic approaches. This section provides details regarding the factor-based approach.

The underlying assumption behind the **factor-based approach** is that a significant portion of a fund's returns can be explained by a set of asset-based factors. This approach involves construction of a portfolio composed of long and/or short positions in a set of suitably selected risk factors that minimize the tracking error with respect to a predefined benchmark. The benchmark may consist of a single manager or, more commonly, an equally weighted benchmark of multiple managers, such as a hedge fund index.

31.5.1 Four Issues in Constructing a Factor-Based Replication Product

The following four issues must be addressed in constructing a replication product using the factor-based approach:

1. Choice of benchmark: The benchmark to be replicated must be selected carefully. The most common practice is to use a publicly available index, such as one of the strategy indices from Hedge Fund Research (HFR) or CISDM. It is also possible to create a custom benchmark that meets certain criteria. A custom benchmark for a strategy (e.g., equity long/short) may be created by ensuring that only those managers who clearly follow such a strategy are added to the benchmark, rather than relying on managers' self-declared strategy.

   When deciding on a suitable benchmark, a decision should also be made on whether to use an investable hedge fund index or a noninvestable hedge fund index for replication. Choosing between investable and noninvestable versions of a hedge fund index involves a trade-off between better replication and targeting a higher return. While tracking products have a much better fit with investable hedge fund indices, as the underlying hedge funds may invest in more liquid securities, noninvestable hedge fund indices often have higher returns than those of investable hedge fund indices, perhaps due to investments in less liquid securities or to the closure of highly successful funds to new investments.
2. Choice of factors: The factors should be readily investable. Otherwise, the replication product will fail to serve its main purpose. In addition, investors should decide whether a fixed set of factors will be used or the most suitable set of factors will be selected using some statistical method (e.g., stepwise regression).
3. Length of estimation period: The parameters of the model have to be estimated using historical performance data. In general, a longer data series means smaller errors in the estimated values of the parameters. However, when there are significant changes in the characteristics of the benchmark and/or the factors, longer estimation periods may lead to a mimicking portfolio that reflects average market conditions over several months rather than current market conditions.
4. Number of factors: Using a large set of factors will generally lead to better in-sample fit but may lead to poor out-of-sample performance. Using a small set of

factors, on the other hand, may prevent the creation of a tracking portfolio with low tracking error, resulting in both poor in-sample fit and poor out-of-sample performance.

Aside from these basic issues, there are many other technical issues to consider, such as adjusting the benchmark's return series for stale prices, adjusting net returns for fees, selecting the appropriate econometric technique to be employed, and using overlays such as volatility control or stop-loss control.

31.5.2 Three Steps to Factor-Based Replication

The following steps are involved in setting up a factor-based replication program once the issues described in the previous section have been resolved.

Step 1: Estimate weights of risky assets

Weights of risky assets can be estimated using the following equation:

$$R_{t,HF} - r_f = \beta_1 \times (F_{1t} - r) + \beta_2 \times (F_{2t} - r) + \cdots + \beta_K \times (F_{Kt} - r) + \varepsilon_t \quad (31.2)$$

where $R_{t,HF}$ is the total rate of return for month t on the benchmark that is being replicated, r_f is the short-term riskless rate (e.g., 30-day T-bill) for month t, F_{it} is the random total rate of return on factor i for month t, β_i is the exposure of the benchmark to factor i, and ε_t is the return on the benchmark that cannot be explained by the combination of factors being used in Equation 31.2. Note that the estimated betas represent the weight of each risky asset in the replicating portfolio. The weight of cash is determined later (see step 2).

The in-sample fit of the model from Equation 31.2 is estimated using r-squared of the regression, which is calculated using the following expression:

$$(1 - Var[\tilde{\varepsilon}]/Var[\tilde{R}_{t,HF}]) \quad (31.3)$$

If the variance of the error term from Equation 31.2 is small, then the in-sample R-squared would be high, indicating a good fit for the model. However, even if the in-sample R-squared is close to 1.0, it does not mean that the out-of-sample tracking error will be small, since most hedge funds actively trade during the month, resulting in changes to allocations for various securities and asset classes. Therefore, the estimated weights are likely to have significant tracking error in out-of-sample periods.

Two important points should be kept in mind while running the regression following Equation 31.2. First, all factors should represent investable assets, as the resulting portfolio needs to be investable. Excess returns on different equity, fixed-income, and commodity ETFs or futures contracts can be used to represent the factors. ETFs and futures trade in liquid markets, provide immediate access to various sources of return, and represent diversified portfolios. Second, returns in excess of a riskless rate must be used on both sides of the equation to relax the requirement that the betas estimated from the equation should add up to 1.0. In this setting, the weight of cash is the free variable and is determined given the weights (i.e., betas) of the risky assets.

Step 2: Estimate the weight of cash

Once the betas (i.e., the weights) for the risky assets are estimated, the weight for cash in the replicating portfolio is given by:

$$\beta_{Cash} = 1 - \sum_{i=1}^{K} \beta_i \tag{31.4}$$

Following this procedure ensures that the weights of all assets, including cash, sum to 1.0. If the weight of cash is negative, then the product will use leverage to create the mimicking portfolio.

Step 3: Invest in different assets

The parameters of the preceding equations are estimated using T observations, with T being the observation for the most recent month; that is, $t = 1, \ldots, T$. The out-of-sample return on the replicating portfolio—the realized return at time $T + 1$—is given by

$$R_{\mathrm{Re},T+1} = \hat{\beta}_{t,T} F_{1,T+1} + \ldots + \hat{\beta}_{K,T} F_{T+1} \tag{31.5}$$

Note that the process is repeated every month as new observations on the performance of the benchmark and the factors become available.

31.5.3 Two Key Concepts Regarding Factor-Based Replication

Two common questions arise in the context of factor-based replication. First, how can hedge fund returns be replicated using a relatively small set of factors? After all, each hedge fund may invest in hundreds of securities, and thus several hundred securities will underlie a hedge fund index. Second, how can the appropriate weights of the mimicking portfolio be estimated if the weights of the actively managed product change over time? If managers are dynamically changing the weights of their portfolios, then the true values of the betas are constantly changing and thus can only be estimated with significant errors. In addition, the estimated betas would reflect what the managers were holding on average over several months, not necessarily what they are holding now.

Two key concepts help to answer these questions.[6] The first concept, referred to as **view commonality**, is related to the fact that when the views of individual hedge fund managers (measured by their exposures) are aggregated in a hedge fund index, they tend to cluster into common themes that drive the overall performance of the index. For instance, if most equity long/short hedge fund managers have positive views on energy stocks, they may attempt to exploit this view by allocating assets to various companies that have exposure to the energy sector. In the index, these views are aggregated and are represented by increased exposure of the index to the energy sector, which could be captured by the replication product through increased allocations to an energy ETF.

The second concept, referred to as **exposure inertia**, asserts that the overall weights of an index consisting of actively managed portfolios can be empirically estimated because the overall exposures change relatively slowly through time. The idea here is that a large number of hedge fund managers underlying a hedge fund index can reduce the speed at which the common views or exposures change over time. If the core themes that drive hedge fund returns change rapidly, then the factor-based replication models would not be able to immediately identify the appropriate weights for the mimicking portfolio. This is particularly true when monthly data are used to estimate the parameters of the model. However, even though one hedge fund may actively change its exposures rapidly, the index is likely to display a more stable behavior because of its exposure to many managers.

31.5.4 Research on Factor-Based Replication

There are many academic and industry research papers on factor-based benchmarking and replication of hedge funds. Schneeweis, Kazemi, and Karavas (2003) examined the replication of equally weighted portfolios of European-based hedge funds for five strategies: fixed-income arbitrage, convertible arbitrage, funds of funds, long/short equity, and event-driven. For this examination, the in-sample period was a 24-month rolling window; the out-of-sample period was the month immediately following the sample window. Selected factors were European market factors providing exposure to equity market risk, interest rate risk, credit risk, and volatility risk. The authors obtained in-sample *r*-squareds of 29.2% for fixed-income arbitrage, 31.9% for convertible arbitrage, 54.3% for funds of funds, 67.7% for long/short equity, and 85.8% for event-driven. The explanatory power of the different factors, by strategy, was low in the in-sample period, generally below 70%. Out-of-sample results were in line with the disappointing in-sample results. The correlation in performance between strategies and clones ranged from 12% to 91% but was mostly below 50%. Only one strategy, event-driven, revealed a satisfying correlation of 90% between the underlying index and the clone. Overall, the replicating models considerably underperformed hedge funds; moreover, they were generally more volatile.

Studying eight strategies, Agarwal and Naik (2004) used a multifactor model in which the risk factors were buy-and-hold and option based. The buy-and-hold risk factors were equities (four indices), bonds (three indices), currencies (one index), and commodities (one index). The authors added the Fama-French size and book-to-market factors, a momentum factor, and a credit risk factor. The option-based risk factors were at-the-money (ATM) and out-of-the-money (OTM) European call and put options on the S&P 500. First, a stepwise regression was used to identify significant factors from 1990 to 2000 for eight HFR indices. Second, the authors examined whether the replicating portfolios based on these factor loadings could do a good job of mimicking the out-of-sample performance of hedge funds. A replicating portfolio was constructed, and the accuracy of replication was tested. The authors obtained in-sample adjusted *r*-squareds ranging from 40.5% to 91.6%, while the one-year out-of-sample results were too short to draw any meaningful conclusion from them.

Hasanhodzic and Lo (2007) attempted to replicate the return distributions of several hedge fund strategies. The exposure of indices representing these strategies to various factors—including returns to asset classes and volatility factors—were estimated. The authors selected those factors that had related liquid instruments, such as forwards and futures contracts. Hasanhodzic and Lo implemented a 24-month rolling window approach, which involved dynamic rebalancing of the portfolio. They reported highly contrasting results. Equally weighted portfolios of rolling window clones underperformed equally weighted portfolios of their respective funds in more than half the strategies. The replication of the correlation with market indices is of particular interest because it can be seen as an indication of the diversification power of hedge funds. The authors analyze the replication of the correlations with various indices. The results, using rolling window linear clones, indicate that returns earned on the clone and an equally weighted portfolio of hedge funds have similar signs in most cases, but could differ by significant amounts in some periods. The results confirm that perfect replication is not possible, and some of the benefits of hedge funds may not carry completely over to replication products.

Lee and Lo (2014) analyze the performance of ASG Global Alternatives Fund, a hedge fund replication product that uses a factor-based replication technique. Based on the performance of the ASG Global Alternatives Fund, the authors conclude that hedge fund beta replication did achieve its objective of providing investors with the liquid portion of the expected returns and the diversification benefits of hedge funds without the complexities and fees of hedge fund investments. In particular, the authors find that the ASG Global Alternatives Fund outperformed the BarclayHedge Fund of Funds Index with lower volatility over the study period. They also report that the correlation between the ASG Global Alternatives Fund and the S&P 500 is close to the correlation between the BarclayHedge Fund of Funds Index and the S&P 500, implying that the diversification benefits of hedge funds are also offered by the clone.

The factor-based models discussed so far represent a small portion of available research focused on explaining hedge fund returns using a factor-based approach. Amenc et al. (2010) use a slightly different variant of a factor-based approach employing nonlinear factors. The idea behind this approach is that hedge fund managers have the skill to time the market and therefore will increase exposure to a factor if the return from that factor is expected to be positive, and decrease exposure to the same factor when the factor return is expected to be negative. The linear factor model of Equation 31.2 can be adjusted in the following manner:

$$R_{t,HF} - r = \beta_1 \times (F_{1t} - r) + \beta_2 \times (F_{2t} - r) + \ldots + \beta_K \times (F_{Kt} - r) + \varepsilon_t \quad (31.6)$$

It can be seen that the last factor assumes that the hedge fund will have an exposure only if factor K is expected to have a positive excess return. Amenc et al. estimate Equation 31.6, along with two other approaches. They report that while the nonlinear models produce a better in-sample fit, their out-of-sample performance is similar to that of the linear model given by Equation 31.2. The authors further attempt to fine-tune their models by selecting factors that are most relevant to each strategy. This results in relatively small improvements to the performances of the models. They conclude that none of the methodologies generate fully satisfactory results, and that the factor-based approach to hedge fund replication faces a series of formidable challenges. These challenges notably include the difficulty of identifying the right factors, as well as the difficulty of replicating, in a robust manner, the time-varying exposures of hedge fund managers with these factors.

Overall, a review of the studies that attempt to replicate hedge fund returns through a factor replication approach leads to the conclusion that replication accuracy is not satisfactory. In-sample r-squared is not sufficiently high to indicate satisfactory in-sample fit, while out-of-sample results suggest that hedge fund return replication is approximate at best.

31.6 PAYOFF-DISTRIBUTION APPROACH

The payoff-distribution approach was developed by Amin and Kat (2003), based on initial theoretical work done by Dybvig (1988) and later applied by Robinson (1998).

31.6.1 Overview of the Payoff-Distribution Approach

The objective of this methodology is far less ambitious than the one pursued in the factor-based approach to hedge fund replication. While the factor-based approach aims to produce a portfolio whose per-period returns match those of the underlying benchmark, the **payoff-distribution approach** aims to produce a return distribution that matches a desired distribution (e.g., that of the benchmark). That is, while this approach was first developed to match the distribution of a hedge fund benchmark, later developments employed this technique to create return distributions that possessed some desirable properties rather than matching a particular hedge fund strategy. For instance, the approach can be employed to create return distributions that are truncated and therefore have limited downside risk. Of course, unless there are certain inefficiencies in the market, creating such a return distribution is likely to be rather costly, and therefore the mean return of the strategy could turn out to be rather low.[7]

31.6.2 Illustration of the Factor-Based and Payoff-Distribution Replication Approaches

The following simple numerical example can highlight the difference between the factor-based replication approach and the payoff-distribution replication approach.

Exhibit 31.3 displays the monthly returns on a hypothetical hedge fund and the returns on two replicators: a factor-based replicator and a payoff-distribution replicator.

Exhibit 31.4 displays some basic statistics related to the performance of these three investment products. The factor-based replicator does a reasonable job of tracking the performance of the hedge fund on a monthly basis. In fact, the correlation between the two is nearly 0.80. However, the factor-based replicator was unable to produce the same mean return, standard deviation, skewness, and kurtosis as those of the hedge fund. On the other hand, the payoff-distribution replicator almost exactly matches the higher moments of the distribution of the hedge fund's return, but does a poor job of tracking the monthly returns on the hedge fund. It can be seen that monthly return from the payoff replicator has very low correlation to the monthly return on the hedge fund. Another important point is that the payoff distribution cannot, and does not try to, match the mean return on the hedge fund. There are two reasons for this.

First, while higher moments (e.g., standard deviation, skewness, and kurtosis) of the distribution of most investments tend to be rather stable, the first moment (i.e., mean) is highly unstable and unpredictable. Thus, it is difficult to create a portfolio

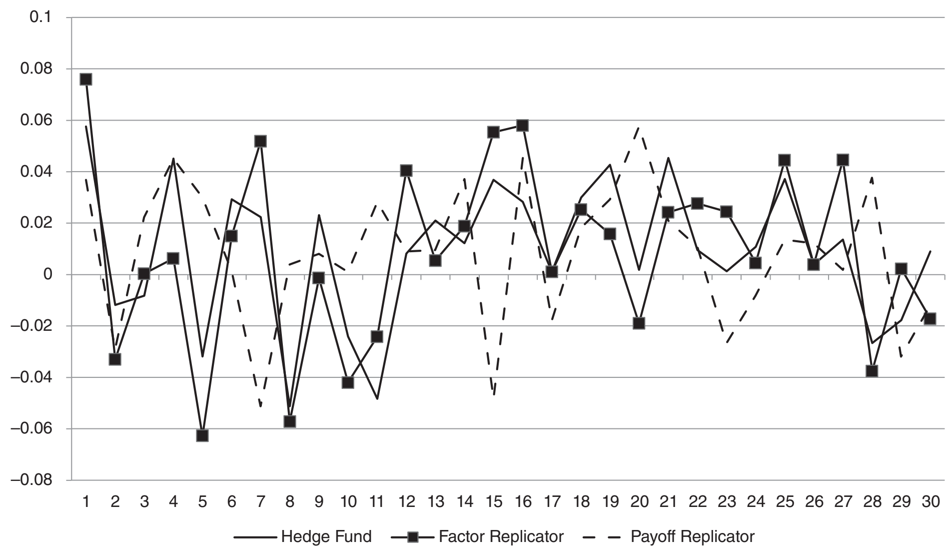


EXHIBIT 31.3 Hypothetical Returns on a Hedge Fund and Two Replicators
Source: Authors' calculations.

that would match the mean return of the benchmark with high probability. The second reason is related to the methodology employed to create the clone. The clone is created using a dynamic strategy similar to the one employed to delta-hedge an option position. When delta hedging is used to replicate an option, one of the assets employed to create the hedged portfolio is the underlying asset of the option. Thus, one is assured that the mean return on the replicator and the option will be closely related. In the case of hedge fund replication, an asset other than the underlying hedge fund has to be used to replicate the hedge fund. Thus, there is no assurance that the mean returns will match. In fact, the higher moments of a hedge fund's return distribution may not match as well if the underlying assets used to replicate the hedge fund have unstable higher moments.

31.6.3 The Key Difference between Payoff-Distribution and Factor-Based Replication

The payoff-distribution approach attempts to replicate the payoff distribution of the hedge fund rather than match individual monthly returns. That is, the factor-based

EXHIBIT 31.4 Statistics of Hypothetical Hedge Fund and Two Replicators

	Hedge Fund	Factor Replicator	Payoff Replicator
Mean	0.90%	0.84%	0.85%
Standard deviation	2.76%	3.46%	2.75%
Skewness	–0.47	–0.219	–0.48
Kurtosis	–0.22	–0.36	–0.22
Correlation vs. hedge fund	100.0%	78.6%	7.4%

and payoff-distribution approaches use the term *replication* in a different sense. The factor-based approach attempts to replicate the time series of monthly returns on a benchmark. The payoff-distribution approach attempts to replicate the end-of-the-period (e.g., one month) distribution of a benchmark.

The idea behind the payoff-distribution approach is rather simple, and is inspired by the option pricing theory of Black and Scholes. More specifically, suppose we desire to replicate the probability distribution of monthly returns on a given hedge fund benchmark. Suppose the monthly returns are denoted by $R_{t,HF}$, and their probability distribution is given by $F(R_{HF})$. This means that

$$\text{Prob}(R_{HF} < x) = F(x) \tag{31.7}$$

The probability distribution function $F(R_{HF})$ is our targeted or desired distribution. Next, suppose that we wish to replicate this targeted distribution using a liquid security called the reserve asset. We denote the monthly returns on this reserve asset by $R_{t,R}$. The goal is now to find a derivative security, $G(R_R)$, written on the reserve asset such that the distribution of monthly returns on this derivative security is equal to the targeted probability distribution. That is,

$$\text{Prob}(R_{HF} < x) = \text{prob}(G(R_R) < x) \tag{31.8}$$

As can be seen from Equation 31.8, the first key issue in applying the payoff-distribution approach is determining the form of the derivative or the payoff function, $G(R_R)$. Next, we need to come up with a trading strategy involving the reserve asset and a risk-free asset (i.e., cash or a cash-equivalent investment) that would replicate the payoff from the derivative.

One of the early criticisms of the payoff-distribution approach was that it ignored a primary reason for investing in hedge funds—namely, its diversification benefits. Thus, it is not enough to match the return distribution of a hedge fund; it is equally important to match its correlations to other assets as well. Kat and Palaro (2006a, 2006b) tackled this problem and proposed a procedure that, in theory, would create a payoff function that not only matches the distribution of a hedge fund's return but also matches its correlation to the investor's current portfolio.

31.6.4 Empirical Evidence on Payoff Distribution

Amenc et al. (2008) performed straightforward tests of the payoff-distribution approach. They attempted to replicate the payoff distributions of 13 hedge fund indices from January 1997 to December 2006. For reserve assets, they used the nearby futures contract on the S&P 500 Index and on the eurodollar over the same period.

The out-of-sample period starts in January 1999 (i.e., they used the first two years of data to calibrate the model). To check the quality of their replication process, they employed various tests. The first one consisted of analyzing the difference between the first four moments of the observed and replicated returns. The results differed depending on the strategy. First, the average returns on the clones were, in most cases, very different from that of the index being replicated. This was to be expected, since the methodology does not aim to replicate the first moment (i.e., mean return) of the hedge fund return distribution. Interestingly, however, the average return obtained for the clone was always lower than that of the index in the out-of-sample period. This result suggests that extreme caution should be used in choosing the reserve asset, and that the performance of the replicating strategy is not robust with respect to the choice of the risky asset involved and the sample period considered.

Second, the volatility values obtained for the clones were relatively close to those obtained for the hedge fund indices. In addition, the authors reported mixed results regarding the equalities of estimated skewness and kurtosis values for the clones and those of the observed returns on the hedge fund indices. Going beyond individual moments, they also tested the equality of the entire distributions. The best replication process was obtained with the short-selling index, and the worst was obtained with the equity market-neutral index. The payoff-distribution approach was able to match the distribution in less than half the cases.

These results imply that even if one is willing to ignore the differences in mean returns, not all hedge fund distributions can be matched with relative satisfaction. Finally, the authors reported the value of the Sharpe ratio and the historical 95% value at risk (VaR). In all cases, the Sharpe ratio was higher for the observed indices, which was to be expected given their higher mean returns. The VaR measure, by contrast, was fairly similar for the clones and the indices.

Another important point is that the good results obtained for at least some hedge fund strategies were displayed only when a long out-of-sample period was used. In some cases, it took close to eight years to obtain payoff distributions that matched the return distribution of the hedge fund indices. For an investor with more limited patience, the payoff-distribution approach to hedge fund replication can lead to disappointment. These results suggest that the payoff-distribution method produces satisfying results only if the investor is willing to wait a long period before assessing the quality of the replication. Finally, the authors show that the replicating portfolio performs poorly in terms of matching the correlation of hedge fund indices with other assets. Investors should also examine their goals for replicating the payoff distribution of hedge funds, as replicating returns with negative skewness and excess kurtosis may not be desirable.

31.7 ALGORITHMIC (BOTTOM-UP) APPROACH

The algorithmic (or bottom-up) approach is substantially different from the previous two approaches, as it is suitable only for well-defined strategies that involve systematic trading and exclude manager discretion as a significant source of return. Examples of relatively well-defined strategies are merger arbitrage, convertible arbitrage, trend following (e.g., CTAs), and a subset of equity long/short (i.e., momentum and value/growth).

31.7.1 Overview of the Algorithmic Approach

The **algorithmic approach** does not rely on a predefined benchmark; it involves implementing a simplified version of the actual trading strategy employed by funds that follow the particular strategy. For instance, the most basic type of merger arbitrage strategy involves taking long positions in target firms and short positions in acquiring firms. Similarly, basic convertible arbitrage involves taking a long position in convertible bonds of a firm and a specific short position in the equity of the same firm such that the equity risk of the convertible bond is hedged away.

To the degree that a systematic trading process representing the fund strategy can be developed and monitored, systematic algorithmic trading strategies can be implemented using a bottom-up approach. The bottom-up approach may not offer some of the benefits of replication, such as liquidity and flexibility, since funds implementing the bottom-up approach essentially hold some of the same securities as those of the hedge funds being replicated. In times of market distress, such as during the most recent financial crisis of 2007–8, bottom-up replication managers faced significant difficulty meeting redemption requests, which, in turn, adversely affected the value of their funds. However, if liquidity risk and event risk are significant contributors to the alpha of hedge fund strategies, investing in hedge funds or bottom-up replication products may be the only way to access this portion of returns.

Next we will look at how to implement a few algorithmic investment strategies.

31.7.2 An Illustration of the Algorithmic Approach: Merger Arbitrage

Merger arbitrage is a strategy that can be implemented by buying a portion of all announced mergers and shorting the acquiring firms. Profitability of this strategy depends on the percentage of mergers that are successfully completed and on the manager's ability to buy shares of the target company at a price below the deal price. Returns are also generated from selling short stock of the acquiring company, which may decline over the course of merger negotiation, and from rebates received on cash generated from selling short the acquirer's stock.

The following hypothetical example will be used to illustrate how merger arbitrage works. Men's Clothing, a men's dress clothes retailer, initiates a hostile takeover bid for Women's Store, another retailer, on November 26, 2015. The initial bid is at a price of $55 in cash for each share of common stock of Women's Store. The day the deal is announced, Women's Store closes at $56.29. On the same day, stock of Men's Clothing closes at $50.60. Subsequently, Men's Clothing increases the bid price initially to $57.50 and finally to $65.00, before completing the acquisition on June 19, 2016. The acquisition process has taken a total of 205 days to complete from the day of initiation of the first bid. On the day the deal is completed, the price of Men's Clothing is $55.67. If a fund manager is implementing a merger arbitrage using this deal, she would buy shares of Women's Store and sell short shares of Men's Clothing after the bid is announced. Exhibit 31.5 describes the profit from merger arbitrage involving acquisition of Women's Store by Men's Clothing assuming the fund manager buys 1,000 shares of Women's Store and sells short 1,000 shares of Men's Clothing. Note that some hedge funds will choose not to hedge a cash merger transaction with a short position in the acquirer's stock.

The main risk of a merger arbitrage strategy stems from mergers not being completed for numerous reasons, such as failure to arrange adequate financing, another bidder emerging in the process and offering a higher price for the same company, and willingness of the target company to stay independent by deploying various anti-takeover measures. If mergers are not completed, an investor stands to lose on both the long and the short side of the trade. Therefore, it is critical to be able to assess the

EXHIBIT 31.5 Profit-and-Loss Statement for the Example Merger Arbitrage Trade

Description		Amount
Gain on Women's Store long position	1,000 × (65.00 – 56.29)	8,710.00
Loss on Men's Clothing short position	−1,000 × (55.67 – 50.60)	−5,070.00
Dividend paid on short position	−1,000 × 0.54	−540.00
Short rebate at 1% rate	1,000 × 50.60 × 1% × 205/365	284.19
Total profit (loss) from strategy		3,384.19
Initial investment	1,000 × 56.29	56,290.00
Return on investment over 205 days	3,384.19/56,290.00	6.01%
Annualized return	$(1 + 6.01\%)^{(365/205)} - 1$	10.95%

probability of the merger being completed in order to run a successful algorithmic merger arbitrage fund.

31.7.3 An Illustration of the Algorithmic Approach: Convertible Arbitrage

Convertible arbitrage is another hedge fund strategy that can be implemented in an algorithmic setting. The simple version of the convertible arbitrage strategy involves purchasing a portfolio of convertible bonds and then hedging out the equity risk of the convertible bond portfolio by using short equity positions in the underlying companies. Essentially, the strategy is equivalent to a combination of a long position in a bond, a long position in a call option, and a short position in shares of stock.

The arbitrager's goal is to purchase a convertible bond with an implicit option that is underpriced (i.e., the implied volatility of the option's underlying stock is lower than the volatility that the arbitrager expects). The arbitrager attempts to profit from the underpriced option while avoiding directional risk by using a delta-neutral hedging strategy. If the underlying equity experiences volatility that exceeds the implied volatility, the strategy earns a profit.

The convertible bond's option has a positive gamma exposure and a negative theta exposure. The strategy is generally successful when profits from hedging activities and the option's positive gamma exceed losses from the time decay of the implicit option. This strategy is usually carried out when the implicit call option of the convertible bond is in-the-money (i.e., the delta of the call option embedded in the convertible bond is greater than 0.5).

For example, assume that a convertible bond is trading at 106% of par; that is, a bond with a $1,000 face value is currently trading at $1,060. Assume that the bond pays semiannual coupons at an annual rate of 5%. This bond converts to 31.5 shares of stock and currently has a delta of 0.602. For demonstration purposes, we will make the delta of our portfolio zero by selling short the appropriate number of shares. Assuming that the current stock price is $28.12, the number of shares that need to be sold short to make the convertible bond delta neutral is 22.69 (1,060 × 0.602/28.12). Therefore, the dollar value of stock held short to create a delta-neutral position is $638.04 (22.69 × 28.12). Now, suppose the stock price moves up by 1%. This will cause the value of the convertible bond to increase by more than 0.602%

EXHIBIT 31.6 Hedging of Convertible Bonds

	Before	After	Change
Long convertible bond	1,060.00	>1,066.38	>6.38
Short stock	−638.04	−644.42	−6.38
Portfolio value	421.96	>421.96	>0

(the delta of the convertible bond) because the positive gamma of the option means that the bond's delta will increase as the stock price increases.

Exhibit 31.6 shows the value of the portfolio before and after a 1% increase in the stock price.

After a 1% increase in the price of the stock, the value of the short position in stock (which has a gamma of zero) becomes –\$644.42 (–638.04 × 1.01). At the same time, the value of the convertible bond becomes greater than \$1,066.38 (greater than 1,060 × 1.00602) because the delta of the convertible increased as the stock price increased. Note that the gain in the value of the convertible bond more than offsets the loss in the short position in the stock. This is what is supposed to happen when the convertible bond is delta-hedged and the convertible has positive gamma. If the stock price had fallen in value there would be a profit to the convertible arbitrage because the delta of the convertible would decrease, thereby reducing the loss from the convertible's price decline to being smaller than the gain from the short stock position.

After any stock price increase, the delta of the convertible bond has increased, so additional shares are sold short to return the position to being delta neutral. Conversely, when the stock price declines, the convertible bond's delta declines and the short position needs to be reduced through the purchase of shares. This hedging activity keeps the strategy delta-neutral. The strategy generates return from the combination of volatility, positive gamma, and periodic trading to restore delta-neutrality. Thus, a primary source of returns is from hedging activity that generates profits from realized volatility in the underlying stock. There is a net profit when the realized volatility exceeds the original implied volatility so that the profits from the convertible's long gamma exposure exceed the losses from the option's time decay.

Convertible bond arbitrage returns are driven by several additional sources. Interest received from the convertible bond and the short stock rebate provide return, while dividend payments and the interest cost of leverage diminish return. Usually, significant leverage is employed in convertible arbitrage strategies to generate returns that are comparable to those earned by other hedge fund strategies. The sources of return for this strategy over a six-month period are illustrated in Exhibit 31.7.

The convertible arbitrage strategy can be further expanded by using interest rate futures to hedge out the interest rate risk and using credit default swaps to hedge out the credit risk of the convertible bond portfolio. This strategy is followed mostly when the underlying option of the convertible bond is close to at-the-money. The strategy will perform well if the implied volatility of the options embedded in the convertible bonds is less than realized volatilities on a consistent basis. As a corporate bond would face interest rate risk, credit risk, and liquidity risk, in addition to other minor risks, so would the convertible bond. To the extent that a convertible arbitrage

EXHIBIT 31.7 Sources of Returns to Convertible Arbitrage

Source	Contribution
Interest earned	2.50%
Short stock rebate	0.10%
Dividend payment	−0.20%
Cost of leverage	−0.80%
Return from hedging activities in excess of time decay	0.40%
Unlevered return in 6 months	2.00%
Unlevered return in 1 year	4.00%
Levered (4×) return in 1 year	16.00%

manager wants to accept risks from various sources, she may hedge out some of the unwarranted risks. Some of the credit risk can be mitigated by a short position in the stock of the underlying company. Liquidity risk is also significant in convertible bonds, as these bonds do not trade often and may have a large bid-ask spread. It is usually hard to hedge liquidity risk, but this can be mitigated to some extent by using listed options.

31.7.4 An Illustration of the Algorithmic Approach: Momentum Strategies

There is substantial academic and industry research on the profitability of momentum (trend-following) strategies, which are discussed in detail in Chapters 25 and 26. A simple version of a momentum strategy involves long positions in a portfolio of stocks, commodities, or currencies that have outperformed over previous periods, and short positions in a portfolio of stocks, commodities, or currencies that have underperformed over previous periods. This strategy has been shown to produce a positive rate of return in the long run. A momentum strategy can be illustrated using a commodity trading adviser (CTA).

CTAs are prime examples of hedge-fund-like strategies that predominantly employ trend-following strategies. CTAs primarily invest in futures contracts across currencies, commodities, interest rates, and stock indices. They may also use options on futures and, in some cases, currency forwards to implement their strategies. Roughly two-thirds of CTA funds follow systematic trend-following strategies, but the systematic managers have the bulk of the money managed by CTAs. CTAs extensively use computer algorithms to find patterns in prices, and can go long and short in any market. They also employ significant leverage to enhance returns.

Let's describe a very simple CTA trading strategy. Suppose a fund manager follows a 20-day and 50-day look-back strategy to trade crude oil. This implies that the manager looks at what the prices of crude oil futures contracts were 20 and 50 days ago, and compares those prices to the current price. If the current price is higher than prices from both 20 days ago and 50 days ago, then the manager opens a long position in a certain number of crude oil futures contracts. The number of contracts is dependent on the risk capital allocated to crude oil. If a large amount of risk capital is allocated to crude oil, a large number of futures contracts are purchased. On the other hand, if the current price is higher than the price from 20 days ago but lower than the price from 50 days ago, the manager does not do anything, since the trend-following strategy is indicating a mixed signal. Risk capital allocated to crude oil may be assigned to other commodities or markets when there is no investment in crude oil. If, however, the current price is lower than prices from both 20 days ago and 50 days ago, then the manager sells short a certain number of crude oil contracts. Thus, the manager can have a long position, a short position, or no position in crude oil at any point in time depending on where current price is relative to price from some time ago. Similarly, the manager can decide what other commodities or currencies or interest rates he wants to trade and can develop trading strategies accordingly.

Volatility of commodities or currencies plays a predominant role in deciding how much risk capital to allocate to each market. If a commodity has higher volatility, a smaller amount of risk capital is allocated to that particular commodity so that the contribution of volatility from that commodity is similar to the volatility from other commodities in the portfolio. This is similar to the approach employed in products based on risk parity.

31.8 ALTERNATIVE MUTUAL FUNDS

Over the past few years, hedge fund strategies have also been offered by mutual funds through a group of funds that have come to be known collectively as alternative mutual funds (AMFs) or liquid alternative funds.

31.8.1 Overview of Alternative Mutual Funds

These funds are known as AMFs because they may invest in futures and options besides stocks and bonds, and they may hold short positions as well as employ leverage. Traditional mutual funds invest only in stocks and bonds, and often do not hold short positions; nor do traditional mutual funds employ leverage. Most AMFs implement underlying hedge fund strategies rather than follow any replication strategies.

AMFs offer daily liquidity similar to that of traditional mutual funds. In order to provide daily liquidity, only a few hedge fund strategies that primarily invest in highly liquid securities can be implemented in a mutual fund structure. Some of the hedge fund strategies that have been successfully implemented in a mutual fund structure are equity long/short, equity market-neutral, managed futures, merger arbitrage, and global macro. There is also restriction on the amount of leverage that can be employed in a mutual fund. As a result, some of the fixed-income arbitrage strategies that require significant leverage to generate meaningful returns are not suitable in a mutual fund structure.

According to a report by Hanouna et al. (2015) from the Securities and Exchange Commission (SEC), AMFs are the fastest-growing segment among all types of mutual funds. Assets under management in these funds grew from $365 million in 2005 to $334 billion in 2014, a nearly 1,000-fold increase in around 10 years. Most of this growth in assets took place between 2010 and 2014, when assets grew from around $60 billion to $334 billion. The number of funds categorized as AMFs also exploded over the same period, growing from fewer than 50 in 2005 to around 600 by 2014.

As the number of funds grew over time, the average assets under management by a fund also grew.

As the number of AMFs has increased, commercial data service providers such as Lipper or Morningstar have begun to categorize these funds into different groups. However, when using commercial mutual fund databases to select AMFs as substitutes for hedge funds, an investor needs to keep at least two things in mind.

First, not all funds categorized as AMFs actually follow hedge fund strategies. For example, a large number of bond funds are categorized under unconstrained fixed income, a subcategory under AMFs. Funds classified as unconstrained fixed income may take advantage of additional flexibility to dynamically adjust exposure to interest rate, currency, inflation, and credit risks so that they can generate risk-adjusted returns in excess of traditional bond indices. These bond funds may not actually be comparable to any hedge fund strategies. In this section, the term *AMF* refers only to those funds under AMFs that follow a hedge fund strategy.

Second, funds categorized under a category—for example, equity long/short—may actually follow an investment strategy that is different from the name of the category. As found in a study by McCarthy (2014), of the 87 funds categorized by Morningstar as equity long/short in January 2013, only 45 actually follow a traditional equity long/short investment strategy and are comparable to hedge funds that follow the same investment strategy. Blindly choosing funds as hedge fund substitutes based on investment strategy assigned by commercial data providers could lead to disastrous consequences.

31.8.2 Three Potential Benefits of Offering Alternative Mutual Funds

With the increase in alternative mutual funds in terms of both assets under management and number of funds over the past 10 years, one must wonder why some hedge fund managers have begun to offer alternative mutual funds that seemingly charge lower fees than those charged by hedge funds. Let us explore some of the benefits that accrue to the fund managers who decide to offer AMFs when they are already operating a hedge fund.

First, a hedge fund manager might be able to raise significantly more capital through an alternative mutual fund. As is well known, most of the new flows to hedge funds are allocated to very large, well-known funds, leaving a large number of small funds with very little chance to raise additional capital. In contrast, mutual funds based in the United States had total assets of $12.7 trillion in 2014, according to Hanouna et al. (2015). This is significantly larger than the pool of assets controlled by all U.S.-based hedge funds. Even if a hedge fund is able to attract a tiny portion of total mutual fund assets, the fund could easily multiply its asset base several-fold. This reason alone can untangle the puzzle of why a lower-fee mutual fund might be offered by a hedge fund manager. With a larger asset base, the fund manager might end up earning more than what he could earn with the performance fee on a smaller hedge fund.

Second, AMFs allow a hedge fund manager to diversify the fund's investor base beyond those who are investing in the hedge fund. Traditional mutual fund investors are retail investors who usually do not qualify for investing in hedge funds based on regulatory standards such as a minimum level of income or net worth. A mutual fund also offers a hedge fund manager the ability to sell the fund through multiple retail channels, enabling it to raise assets from a diverse set of investors.

Third, increased regulations have closed the gap between hedge funds and mutual funds. In the case of the United States, with the implementation of the Dodd-Frank Act (2010), many hedge fund managers, albeit reluctantly, had to register as investment advisers. Since a U.S. mutual fund manager must also be registered as an investment adviser, a hedge fund manager can easily launch a U.S. mutual fund once registered. In addition, having a mutual fund removes certain restrictions that are placed on a hedge fund. For example, a mutual fund can attract as many investors as are willing to invest in the fund, whereas a hedge fund may face restrictions on the number of investors that it can accommodate.

Not only do fund managers benefit from AMFs, but investors also benefit from having access to alternative mutual funds. The following section offers an analysis of benefits of AMFs to retail investors.

31.8.3 Benefits of Alternative Mutual Funds to Investors

Besides offering hedge fund strategies in a mutual fund structure, the biggest benefits of AMFs for investors are transparency, daily liquidity, and low fees, while operating in a regulated framework that offers various safeguards to both retail and institutional investors. In order to provide daily liquidity, U.S. AMFs must limit investments in illiquid securities to 15% of total assets under management. These AMFs must also provide holding-level transparency through quarterly filings of holdings to the Securities and Exchange Commission. Since all U.S. mutual funds must comply with the U.S. Investment Company Act of 1940 (the '40 Act), they must also limit their leverage to 33% of the gross value of the fund, maintain sufficient diversification by limiting investment in a single security to not more than 25% of the fund's portfolio, not charge an incentive fee, and derive at least 90% of their income from permitted sources. Each of these requirements can have both favorable and unfavorable effects on the risk-return characteristics of any alternative mutual fund. On the beneficial side, operational risks of the alternative mutual fund may be lower than those of a similar hedge fund that operates in an unregulated environment, reducing possibilities of losses due to fraud. The downside of such requirements is that expected returns of the alternative mutual fund would be lower than those of a hedge fund with a similar strategy.

One of the biggest benefits of alternative mutual funds is the lack of minimum eligibility requirement for investing in them. In order to invest in a hedge fund, a U.S. investor must be either an accredited investor (net worth of $1 million or more, excluding the value of the primary residence, or two consecutive years of a personal income of $200,000 or a household income of $300,000) or a qualified purchaser (net worth of $5 million or more). No such requirements need to be met when someone wants to invest in an alternative mutual fund. There are many investors who would not be able to access hedge fund–like strategies if alternative mutual funds were not available.

Another key advantage of alternative mutual funds is their lower fees compared to hedge funds. Whereas hedge funds usually charge an incentive fee in addition to a hefty management fee of around 2% of assets under management, mutual funds cannot charge an asymmetric incentive fee, and their management fees are usually lower than 2%. The cumulative effect of high fees on investment performance is significant, and the underlying fund would have to earn a lot more if after-fee performance is considered, as was discussed earlier in this chapter.

Having a hedge fund–like strategy in a portfolio offers the benefit of diversification for most retail investors who hold stock and bond funds. At times of market stress, diversification can protect a portfolio from having large drawdowns and can improve risk-adjusted return.

31.8.4 Risks of Alternative Mutual Funds

Even with all the benefits listed for both investors and fund managers, AMFs have some risks that one needs to study carefully before investing. This section discusses a few of these risks.

Perhaps the biggest risk for an AMF is liquidity. Even with the provision of not investing more than 15% of assets in illiquid assets, funds may at times find themselves facing unforeseen redemption requests, which might require them to sell assets that have low liquidity. A surge in redemption requests usually depletes the most liquid assets first and subsequently requires the selling of assets that are less liquid. This results in two unfavorable conditions for the fund. First, as the fund starts selling less liquid assets at depressed prices, the NAV of the fund suffers. This leads to an exodus of investors who want to leave the fund before others, exacerbating an already dire situation. Second, as funds start selling more liquid assets first, the remaining investors are saddled with less liquid assets. This again leads investors to want to sell out of their shares before others, resulting in a deadly spiral that ends badly for all. If, on the other hand, AMFs invest in more liquid assets to meet unexpected large redemption requests in a timely manner, they may only be able to offer a watered-down version of a hedge fund, which may not offer all the diversification benefits investors have come to expect from hedge fund–like strategies.

Another risk for AMFs originates in the excessive leverage inherent in different securities they may invest in. Even though AMFs are not permitted to have more than 33% leverage, they may invest in securities that are themselves leveraged. For example, AMFs may invest in levered ETFs or futures contracts, which have high leverage that is not accounted for if one is looking at only the amount borrowed by a fund. Leverage has both advantages and disadvantages. It offers high returns when markets move in a fund's favor, but brings disastrous effects when markets move against a fund. Significant attention needs to be paid in analyzing portfolio holdings and investment policies of an AMF before investing.

Another risk that is more operational in nature is related to trade allocation when the same investment manager operates a hedge fund and an AMF side by side. As hedge funds earn incentive fees in addition to higher management fees, the temptation is great for a fund manager to allocate favorable trades to the hedge fund to maximize his personal benefit at the expense of AMF investors.

31.9 EXCHANGE-TRADED FUNDS

Some hedge fund strategies have also been offered through exchange-traded funds (ETFs) since 2009. ETFs combine many of the features of mutual funds with the trading features of common stocks. ETFs offer three distinct advantages over mutual funds. First, ETFs can be bought or sold throughout the trading day, making it easy to get into or out of any position. Second, ETFs disclose their holdings at the start of every trading day, unlike mutual funds, which disclose holdings once a quarter. This not only makes it easy for investors to assess fair value of an ETF share, but it also prevents an ETF from drifting away from its stated objective, as might happen to a mutual fund. If an ETF share deviates too much from the value of underlying stocks held in one share of the ETF, market makers or other authorized agents can take advantage of arbitrage opportunities. Third, ETFs offer tax advantages over mutual funds because they have lower turnover and they allow for in-kind redemptions. In-kind redemption allows an ETF to deliver a group of securities that have appreciated in value in exchange for shares of the ETF, thereby avoiding the taxable event that would have occurred if those appreciated securities had been sold in the open market.

ETFs as a group have garnered a large amount of assets over the past two decades. According to Hill, Nadig, and Hougan (2015), as of the end of Q1 2014, total assets under management by all ETFs in the United States stood at $1.74 trillion. Most of these assets, however, were managed by traditional equity and fixed-income ETFs that follow well-defined passive benchmarks, such as the S&P 500 Index and the Barclays Capital U.S. Aggregate Bond Index. ETFs categorized as alternatives, which include those that follow hedge fund strategies, are only a tiny part of the overall ETF market. Alternatives accounted for 0.18% of total assets managed by ETFs as of March 31, 2014. Some of the major selling points of ETFs, such as low cost and transparency, are lost when one tries to manage an alternative strategy in an ETF structure. Among all the groups of ETFs, alternatives have the highest cost according to Hill et al. (2015).

The largest ETF in terms of market value that follows a hedge fund strategy is IQ Hedge Multi-Strategy Tracker ETF (QAI), which had just over $1 billion in assets as of October 31, 2015. The underlying strategy of this ETF is hedge fund replication through investment in other ETFs. In a sense, this ETF is an ETF of ETFs, just as a fund of hedge funds is a hedge fund that invests in other hedge funds. There are only two other ETFs that had more than $100 million in assets on October 31, 2015, and that follow a hedge fund strategy. One of them follows a managed futures strategy, while the other follows a merger arbitrage strategy.

31.10 CONCLUSION

Hedge fund replication products are based on three different approaches: factor-based, payoff-distribution, and algorithmic (bottom-up) replication. The factor-based approach attempts to take advantage of the fact that beta exposures of hedge funds have increased in recent years, and therefore a larger portion of their returns can be explained by traditional sources of risk and return. In addition, financial innovations in recent years have given investors expanded access to alternative sources of return. Consequently, carefully constructed portfolios of liquid securities may be able to capture a large portion of the returns earned by a portfolio of hedge funds. The payoff-distribution approach attempts to match the return properties of a hedge fund by implementing dynamic trading strategies similar to those employed in a traditional option-replication strategy. Finally, the algorithmic approach attempts to implement a simple and transparent version of a well-defined hedge fund strategy. For example, a simple version of the merger arbitrage strategy would require an investor to take a long position in a portfolio of target companies and a short position in a portfolio of the acquiring companies.

NOTES

1. See Fung and Hsieh (2007); Fung et al. (2008); Naik, Ramadorai, and Stromqvist (2007); Billio, Frattarolo, and Pelizzon (2014); Getmansky, Lee, and Lo (2015).
2. See Jagannathan, Malakhov, and Novikov (2012).
3. Brown, Goetzmann, and Ibbotson (1999); Bares, Gibson, and Gyger (2003).
4. See www.hedgebay.com.
5. Kamel (2007).
6. Drachman and Little (2010).
7. For an application of this approach in creating portfolios with low downside risk, see Hocquard, Papageorgiou, and Remillard (2015).

REFERENCES

Agarwal, V., and N. Y. Naik. 2004. "Risks and Portfolio Decisions Involving Hedge Funds." *Review of Financial Studies* 17 (1): 63–98.

Amenc, N., W. Géhin, L. Martellini, and J.-C. Meyfredi. 2008. "Passive Hedge Fund Replication: A Critical Assessment of Existing Techniques." *Journal of Alternative Investments* 11 (2): 69–83.

Amenc, N., L. Martellini, J.-C. Meyfredi, and V. Ziemann. 2010. "Passive Hedge Fund Replication: Beyond the Linear Case." *European Financial Management* 16 (2): 191–210.

Amin, G. S., and H. M. Kat. 2003. "Hedge Fund Performance 1990–2000: Do the 'Money Machines' Really Add Value?" *Journal of Financial and Quantitative Analysis* 38 (2): 252–74.

Bares, P., R. Gibson, and S. Gyger. 2003. "Performance in the Hedge Fund Industry: An Analysis of Short and Long-Term Persistence." *Journal of Alternative Investments* 6 (3): 25–41.

Billio, M., L. Frattarolo, and L. Pelizzon. 2014. "A Time-Varying Performance Evaluation of Hedge Fund Strategies through Aggregation." *Bankers, Markets & Investors* 129 (March–April): 40–58.

Bodie, Z., A. Kane, and A. Marcus. 2010. *Investments*. 9th ed. Berkshire, UK: McGraw-Hill/Irwin.

Brown, S. J., W. N. Goetzmann, and R. G. Ibbotson. 1999. "Offshore Hedge Funds: Survival and Performance, 1989–1995." *Journal of Business* 72 (1): 91–118.

Crowder, G., H. Kazemi, and T. Schneeweis. 2011. "Asset Class and Strategy Investment Tracking Based Approaches." *Journal of Alternative Investments* 13 (3): 81–101.

Drachman, J., and P. Little. 2010. "Enhancing Liquidity in Alternative Portfolios." Credit Suisse Asset Management.

Dybvig, P. H. 1988. "Inefficient Dynamic Portfolio Strategies or How to Throw Away a Million Dollars in the Stock Market." *Review of Financial Studies* 1 (1): 67–88.

Fung, W., and D. A. Hsieh. 2007. "Will Hedge Funds Regress towards Index-Like Products?" *Journal of Investment Management* 5 (2): 46–65.

Fung, W., D. A. Hsieh, N. Y. Naik, and T. Ramadorai. 2008. "Hedge Funds: Performance, Risk, and Capital Formation." *Journal of Finance* 63 (4): 1777–1803.

Getmansky, M., P. A. Lee, and A. W. Lo. 2015. "Hedge Funds: A Dynamic Industry in Transition." *Annual Review of Financial Economics* 7: 483–577.

Hanouna, P., J. Novak, T. Riley, and C. Stahel. 2015. "Liquidity and Flows of US Mutual Funds." Division of Economic and Risk Analysis, Securities and Exchange Commission.

Hasanhodzic, J., and A. W. Lo. 2007. "Can Hedge-Fund Returns Be Replicated? The Linear Case." *Journal of Investment Management* 5 (2): 5–45.

Hill, J. M., D. Nadig, and M. Hougan. 2015. "A Comprehensive Guide to Exchange Traded Funds (ETFs)." CFA Institute Research Foundation.

Hocquard, A., N. Papageorgiou, and B. Remillard. 2015. "The Payoff Distribution Model: An Application to Dynamic Portfolio Insurance." *Quantitative Finance* 15 (2): 299–312.

Jagannathan, R., A. Malakhov, and D. Novikov. 2012. "Do Hot Hands Exist among Hedge Fund Managers? An Empirical Evaluation." *Journal of Finance* 65 (1): 217–55.

Kamel, T. 2007. "Hedge Fund Replication." Iluka Hedge Fund Consulting. www.ilukacg.com.

Kat, H. M., and H. P. Palaro. 2006a. "Hedge Fund Indexation the FundCreator Way: Efficient Hedge Fund Indexation without Hedge Funds." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=949057.

———. 2006b. "Replication and Evaluation of Fund of Hedge Funds Returns." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=873465.

Lee, P. A., and A. W. Lo. 2014. "Hedge Fund Beta Replication: A Five-Year Retrospective." *Journal of Investment Management* 12 (3): 5–18.

McCarthy, D. 2014. "Hedge Funds versus Hedge Mutual Funds: An Examination of Equity Long/Short Funds." *Journal of Alternative Investments* 16 (3): 6–24.

Naik, N. Y., T. Ramadorai, and M. Stromqvist. 2007. "Capacity Constraint and Hedge Fund Strategy Returns." *European Financial Management* 13 (2): 239–56.

Robinson, B. 1998. "Efficiency Cost of Guaranteed Investment Products." *Journal of Derivatives* 6 (1): 25–37.

Schneeweis, T., H. Kazemi, and V. Karavas. 2003. "EUREX Derivative Products in Alternative Investments: The Case for Hedge Funds." EUREX.

CHAPTER 32

Funds of Hedge Funds and Multistrategy Funds

At the close of 2014, Hedge Fund Research estimated that the hedge fund industry managed a total of \$2,845.1 billion in 10,101 funds. Single-manager hedge funds controlled \$2,171.8 billion, while funds of hedge funds (FoFs) managed a total of \$673.3 billion. FoFs are significant allocators to hedge funds, as they represent 23.7% of all hedge fund assets under management (AUM). Exhibit 32.1 shows that investors have continued to allocate to hedge funds, even after a 25% decline in assets in 2008.

Exhibit 32.2 shows the remarkable growth in the number of hedge funds and FoFs since 1990, when the industry consisted of 530 hedge funds and 80 FoFs, through the end of 2014, when the industry had 8,377 hedge funds and 1,724 FoFs. The previous peak in the number of hedge funds was 7,634 in 2007, a number that has continued to grow to its current historical high. For FoFs, the previous peak was also in 2007, when there were 2,462 FoFs, after which the number gradually declined to its current level.

32.1 APPROACHES TO ACCESSING HEDGE FUNDS

In order to benefit from an allocation to hedge funds, investors must build a diversified portfolio that includes exposure to different strategies and different funds. There are three basic approaches for investors to gain hedge fund exposure in their portfolios: direct, delegated, and indexed.

32.1.1 The Direct Approach and Its Three Advantages

The direct approach consists of investing directly in a number of single-manager hedge funds to create a portfolio. With more than 8,377 hedge funds globally from which to choose, this approach requires extensive resources to identify and research managers, as well as the relevant experience and expertise to determine the appropriate blend of strategies, styles (how each manager executes the same strategy), and funds.

In spite of these obstacles, institutional investors are increasingly choosing to invest directly in hedge funds for a variety of reasons. The three major reasons are:

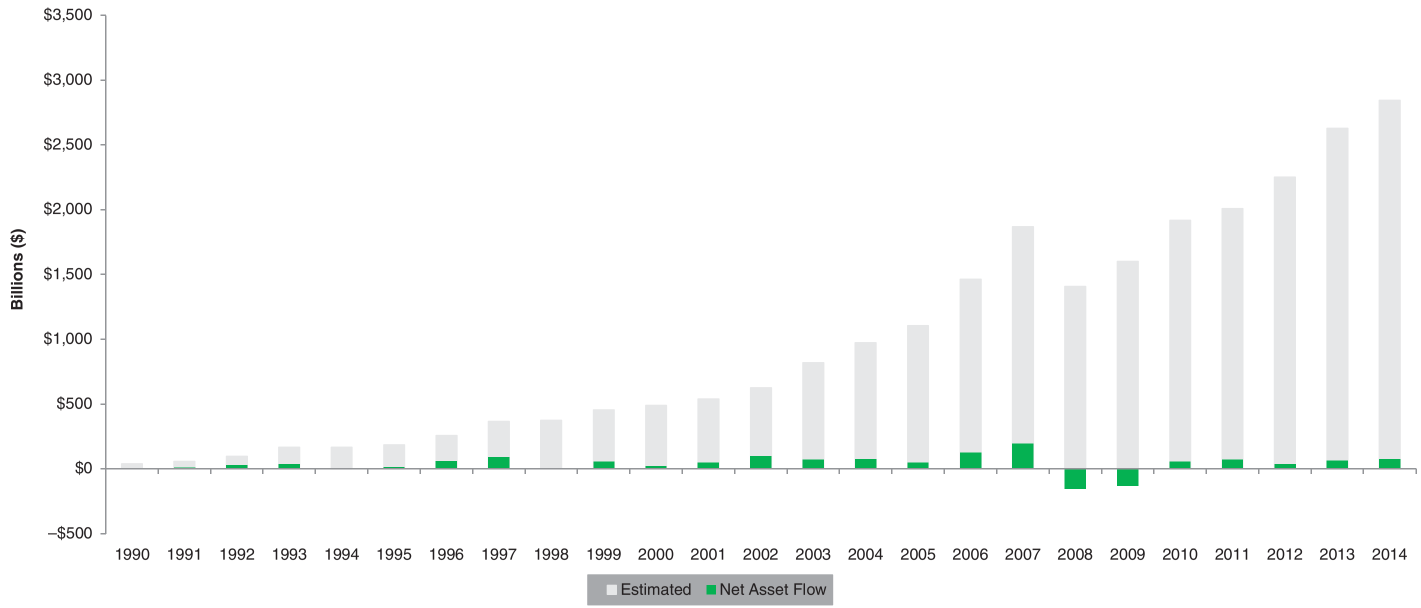


EXHIBIT 32.1 Estimated Growth of Assets and Net Asset Flows for Hedge Fund Industry, 1990–2014
Source: HFR Global Hedge Fund Industry Report (2014).

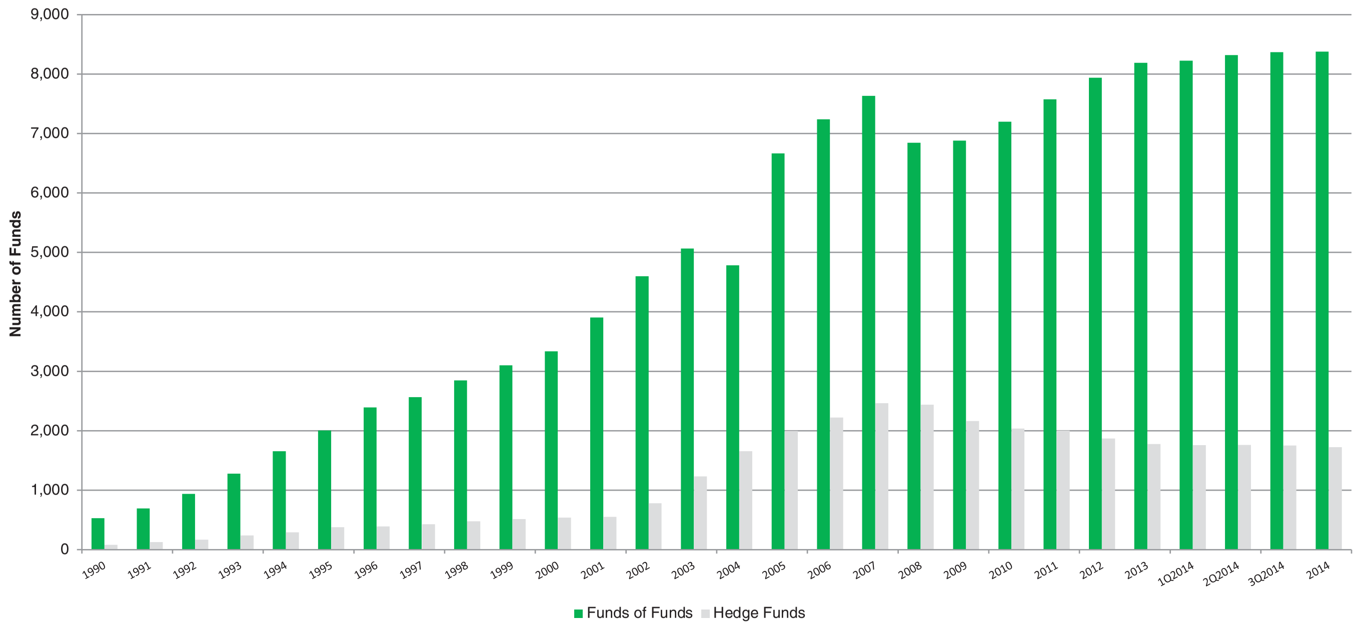


EXHIBIT 32.2 Estimated Growth of the Number of Funds: Funds of Funds versus Hedge Funds, 1990–2014
Source: HFR Global Hedge Fund Industry Report (2014).

1. The cost savings from the extra layer of fees charged by an FoF
2. Access to cost-effective, experienced consultants
3. The ability to have improved control and transparency in the asset allocation and due diligence process

This approach may, however, be challenging for those investors who are constrained by the minimum capital levels and sophistication standards required by regulators in many countries. As a result, small or medium-sized investors may not be able to create a fully diversified portfolio of hedge funds while keeping their total allocations at a reasonable level.

32.1.2 The Delegated Approach and Its Services

In an FoF investment, portfolio management is delegated to another party. This delegated approach provides the following five services:

1. Sourcing managers: As significant allocators of capital, FoFs are actively sought out by hedge funds as clients. Access to higher-quality hedge funds is typically based on the size and quality of the FoF, established relationships with the hedge funds, and the quality of the FoF's client base.
2. Due diligence: Due diligence is the process of understanding, reviewing, and analyzing the strategy, management, and operations of a hedge fund manager. This is perhaps one of the most important value propositions that differentiate an FoF manager for investors who are deciding between a direct and a delegated hedge fund investment program. Unfortunately, several large FoFs have been marred by underlying manager blowups and frauds that have caused some institutional investors to reassess the value of an FoF's due diligence process.

   There is, however, some academic evidence justifying the payment of an additional layer of fees in return for more robust operational due diligence. Operational due diligence is the process of evaluating the policies, procedures, and internal controls of an asset management organization. Brown, Fraser, and Liang (2008) estimate that, net of fees, the largest FoFs outperformed the smallest FoFs by a statistically significant +2.69% per year from 1995 to 2006, and that Sharpe ratios for large FoFs were twice those of smaller FoFs. Larger FoFs may outperform because their scale allows them to invest more in due diligence and risk management processes. Small FoFs are at a competitive disadvantage because, due to their small size, they are not able to generate sufficient fees to cover the fixed and necessary costs of initial due diligence.
3. Strategy and fund selection: The FoF is responsible for selecting appropriate strategies and a proper mix of funds for each strategy. While FoFs may have access to preferred hedge funds as well as experienced insight on which strategies are likely to outperform going forward, institutional investors are catching up. Historically, institutional investors have also used FoFs to reduce the risk of negative headlines should one of the underlying hedge fund investments blow up. The dissemination of hedge fund knowledge and expertise across the industry has minimized this value proposition of FoFs.
4. Portfolio construction: Once the strategies and funds have been selected, the FoF has to decide on position sizing: that is, how much to allocate to each

strategy and each fund to build the portfolio. The allocation will depend on the specific objectives of the FoF in combination with the terms of the FoF offering, including the strategies permitted and liquidity provisions for investors. The portfolio construction process includes evaluating the expected correlations among the underlying hedge funds, ensuring that the FoF is adequately diversified by strategy and style, and making sure that the combined liquidity provisions of the underlying hedge funds conform to the terms offered by the FoF.

5. Risk management and monitoring: The FoF manager will monitor each underlying hedge fund to ensure that its performance profile is consistent with the fund's and the FoF's overall objectives. Some FoFs employ sophisticated risk management processes to monitor the underlying hedge funds' positions, while others may employ multifactor sensitivity analysis to gauge risk exposure to various market factors and to analyze a fund's potential risk.

32.1.3 The Indexed Approach

Investing in index-type products allows an investor to select a representative hedge fund index and buy a financial product that aims to replicate the performance of that index. Indexed financial products are typically sold as certificates or principal-guaranteed notes issued by a credit-worthy financial institution, whose returns are typically linked to the performance of a referenced FoF or specific hedge fund index.

Most recently, other products have proliferated that claim to provide hedge fund exposure without some of the major drawbacks of hedge funds; advantages include lower fees, better liquidity, and improved transparency. Liquid alternative funds employ constrained hedge fund strategies that are offered in a mutual fund structure with daily liquidity. These funds are constrained in terms of the liquidity of the underlying securities in which they can invest, the amount of leverage they can employ, and their use of short positions. Replication strategies, which come in many variations, typically identify market factors to which hedge funds are exposed and may replicate the return pattern through the use of derivatives.

Exhibit 32.3 shows the tremendous growth of the FoF industry, increasing AUM from $1.9 billion in 1990 to $673.3 billion at the end of 2014. FoF assets peaked in 2007 at $798.6 billion as institutions made their initial hedge fund allocations via FoFs. During the financial crisis of 2008–9, FoF assets dropped to a low of $571.3 billion due to a combination of negative performance, withdrawals on the heels of the Madoff fraud, and institutional investors investing directly in hedge funds.

32.2 CHARACTERISTICS OF FUNDS OF HEDGE FUNDS

A fund of hedge funds (FoF) is an investment vehicle that pools the capital from a number of investors and allocates it to underlying hedge funds to achieve a stated objective. By investing in an FoF, investors delegate the management of their portfolios to the FoF manager. The FoF manager is responsible for sourcing individual hedge funds, performing comprehensive due diligence on each manager and fund, securing capacity, managing the overall portfolio, monitoring risks and returns, and reporting performance and other information to investors.

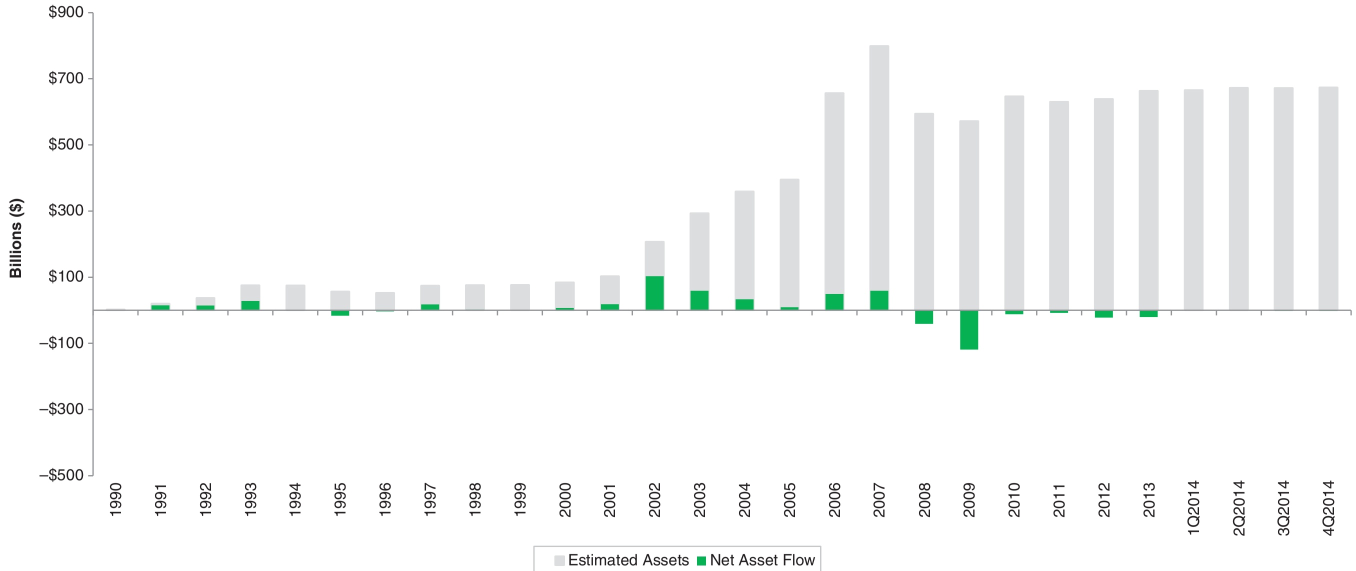


EXHIBIT 32.3 Estimated Growth of Funds of Funds, 1990–2014
Source: HFR Global Hedge Fund Industry Report (2014).

32.2.1 Background on Funds of Hedge Funds

Historically, FoFs charged a 1% management fee and a 10% incentive fee in addition to the management and incentive fees charged by the underlying managers. Disappointing FoF performance during the crisis of 2008, coupled with a strong push by large institutional investors to invest in hedge funds directly, caused fees for institutional investors in FoFs to compress meaningfully. For large allocations, such as those over $75 million, FoFs now typically charge 50 to 100 basis points with no incentive fees. FoFs targeted to high-net-worth clients have not experienced as much fee compression. Typical lockup periods for FoFs depend on the FoF strategy, with liquidity provisions usually mirroring those of the underlying investments. For a diversified FoF, the typical lockup period is one year, while redemptions are quarterly, with notice periods ranging between 90 and 180 days. Subscriptions are generally offered monthly.

The success of FoFs was initially fueled by smaller wealthy investors who wanted to access hedge funds but did not have the capital levels or the expertise to invest directly in hedge funds. In the mid-2000s, FoFs grew out of demand by institutional investors that were new to the alternative investment industry. Institutional investors are generally made up of pension funds, sovereign wealth funds, endowments and foundations, private banks, and family offices. These investors preferred an FoF for their first allocation as a way to learn about the industry and how hedge funds operate. As these investors gained experience, they leveraged their relationships with the FoFs and the underlying managers to make direct investments, moving the FoF closer to the role of consultant rather than asset manager.

Recent asset flows into the hedge fund industry have been decidedly skewed toward larger managers, as institutions have made sizable investments. Many investors place a maximum limit on how much of a manager's AUM they can represent; for example, a $100 million investment would require a fund to have $1 billion AUM if the investor had instituted a limit of 10%. At the end of 2014, the largest 10% of hedge funds (those with assets exceeding $500 million) managed 88% of industry AUM (see Exhibit 32.4). As a result, many of the smaller funds are at a distinct disadvantage in their ability to attract and/or retain adequate AUM levels to support increasingly stringent compliance and regulatory infrastructure.

According to Xiong et al. (2009), FoF performance, fund flows, and asset size are positively correlated. Not surprisingly, the authors found that FoFs that have better performance experience greater capital inflows, whereas the worst-performing FoFs experience net capital outflows. Moreover, they concluded that 18-month Sharpe ratios have more explanatory power for capital inflows than do Sharpe ratios measured over other durations or average raw returns. These findings seem to suggest that small funds with poor performance struggle to survive.

32.2.2 Diversification Types of Funds of Hedge Funds

An FoF's approach to diversification is also a distinguishing characteristic in the universe of FoFs. Broadly speaking, FoFs can be grouped based on the number of underlying strategies, the number of underlying hedge funds, or a stated objective:

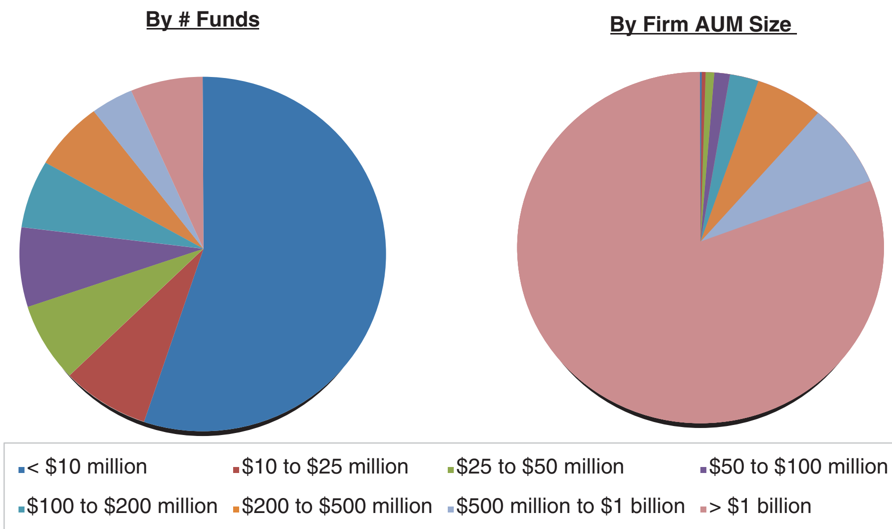


EXHIBIT 32.4 Estimated Distribution of Industry Assets by Fund AUM, Q4 2014
Source: HFR Global Hedge Fund Industry Report (2014).

- Diversified FoFs attempt to diversify a portfolio by allocating assets to a larger number of hedge funds (typically 30 to 50) that follow different strategies and are generally expected to have returns that have low correlations with one another.
- **Concentrated FoFs** typically allocate assets to a relatively small number of hedge funds compared to diversified FoFs (typically 5 to 10).
- **Single-strategy FoFs** allocate assets across several hedge funds (typically 5 to 15) following the same strategy, theme, or group of strategies. Their goal is to provide exposure to a particular subset of the hedge fund universe. Examples of single-strategy FoFs include funds that allocate exclusively to equity hedge funds, managed futures funds, or emerging managers.
- **Tactical FoFs** invest in a group of hedge funds (typically 5 to 10) to opportunistically gain exposure to a specific market factor. A recent example of this was the proliferation of credit-based strategic FoFs in 2009 to profit from the credit dislocation of 2008.

32.2.3 Funds of Hedge Funds Performance Reporting and Potential Biases

FoFs voluntarily report assets, performance, and other fund information to commercial databases that track them. These service providers greatly facilitate information gathering and performance comparisons among the many funds. Moreover, many experts feel that the data on returns measured by FoF databases are usually of a much better quality than the data on individual hedge funds. In particular, some of the usual hedge fund biases are significantly reduced or even eliminated when applied to FoFs (see Fung and Hsieh 2002), as indicated by the following four levels of potential bias:

1. Little or no hedge fund survivorship bias: FoFs provide audited track records, which include full historical allocations to all winners and losers, regardless of whether the underlying funds voluntarily reported to a database. As a result, the historical track record of a hedge fund that stops reporting to databases will be included in the track record of all FoFs that have invested in that fund and reported their performance.
2. No hedge fund selection bias: While hedge funds may choose not to report to databases, their track record will be embedded in the performance of any FoF that was invested in it. As a result, looking at FoF databases increases the potential universe of funds captured to include those that do not report to databases.
3. No instant history bias: When an FoF adds a new hedge fund to its portfolio, the historical track record of that hedge fund is not included in the historical track record of the FoF. As a result, returns from FoFs are less susceptible to instant history bias.
4. Less survivorship bias: The mortality rate of FoFs is much lower than that of individual hedge funds. The performance differential between surviving and liquidated FoFs is also smaller than that of individual hedge funds. According to Liang (2003), the annual survivorship bias for hedge funds is +2.32% per year, while the bias for FoFs is +1.18%.

As a result, FoF returns are more likely to deliver a more accurate estimate of the actual investment experience of hedge fund investors than are single hedge fund returns or hedge fund indices.

32.3 FUNDS OF FUNDS PERFORMANCE

From January 1990 to December 2014, FoF indices have reported average annualized returns between 6.54% and 7.68%, with annual standard deviations between 7.89% and 10.13%, as proxied by the following indices: HFRI FOF: Conservative, HFRI FOF: Diversified, HFRI FOF: Composite, and CISDM FOF: Diversified (see Exhibit 32.5 and Exhibit 32.6). The risk-free rate for calculating the Sharpe ratio is one-month LIBOR.

As shown in Exhibit 32.7, the best years for FoFs were 1993 and 1999, with returns ranging from 16.3% to 28.5%. Clearly the worst year was 2008, with FoFs down between –16.8% and –21.4%.

The period from 1990 to 1999 was much more profitable for FoFs than was the period from 2000 to 2014 (see Exhibit 32.8). FoF average annual returns fell from the range of 10.7% to 13.0% in the period from 1990 to 1999 to the range of 3.73% to 4.20% in the period from 2000 to 2014. More than 3% of this return decline may be explained by a declining interest rate environment, as one-month LIBOR averaged 5.4% in the 1990s and less than 2.17% in the recent period. Similarly, U.S. equity returns declined from an average of 19% to just 6.09%, even though FoFs had a higher average beta in the more recent period than in the prior period. This means that since 2000, FoFs have reported annual returns that are about one-third of the annual returns they reported in the 1990s. This is in line with the sharp drop-off in hedge fund annual returns for the same two periods. Additionally, Sharpe ratios for FoFs fell from the range of 0.62 to 0.76 to the range of 0.21 to 0.31,

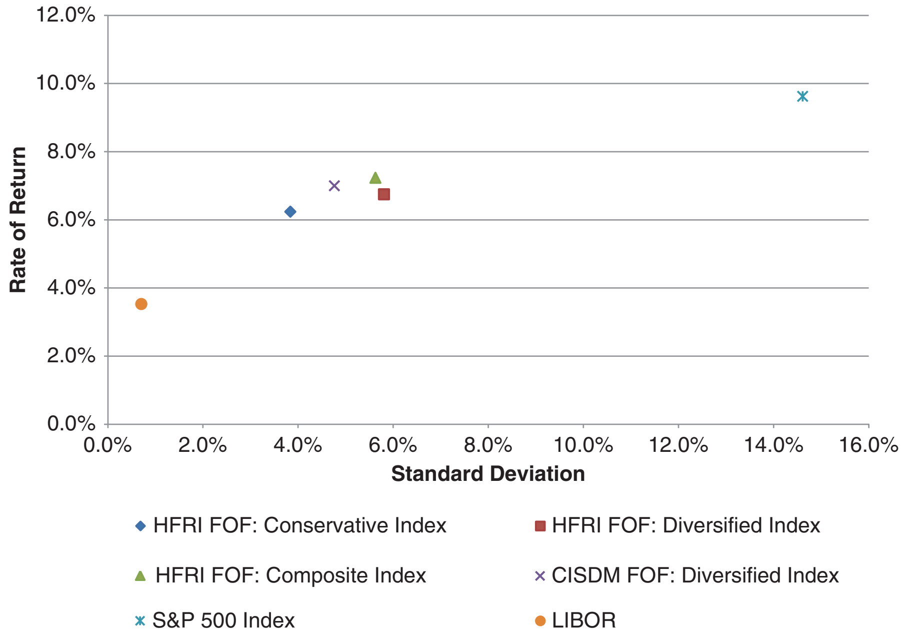


EXHIBIT 32.5 Risk-Return Trade-Off for Funds of Funds, 1990–2014
Source: Hedge Fund Research, Bloomberg, Center for International Securities and Derivatives Markets (CISDM).

which is consistent with the drop in Sharpe ratios reported by hedge funds. As hedge funds have evolved over time, it appears that their role in a portfolio has evolved as well. Early on, hedge funds attracted high-net-worth investors because of their strong performance; most recently, institutional investors have used their hedge fund allocations to reduce the volatility of their aggregate portfolios.

EXHIBIT 32.6 Summary Statistics for Funds of Funds, 1990–2014

1990–2014	HFRI FOF: Conservative Index	HFRI FOF: Diversified Index	HFRI FOF: Composite Index	CISDM FOF: Diversified Index	S&P 500 Index	LIBOR
Annual return	6.54%	7.25%	7.68%	7.32%	11.25%	3.45%
Annual standard deviation	7.89%	9.93%	10.13%	8.32%	18.25%	2.44%
Excess return	3.09%	3.80%	4.23%	3.87%	7.81%	
Beta	0.26	0.28	0.39	0.26		
Alpha	1.06%	1.63%	1.94%	1.82%		
Sharpe ratio (R_f = LIBOR)	0.39	0.38	0.42	0.47	0.43	

Source: Hedge Fund Research, Bloomberg, CISDM.

EXHIBIT 32.7 Fund of Funds Index Performance by Year, 1990–2014

Year	HFRI FOF: Conservative Index	HFRI FOF: Diversified Index	HFRI FOF: Composite Index	CISDM FOF: Diversified Index	S&P 500 Index	LIBOR
1990	14.2%	17.0%	17.5%	7.5%	–3.1%	8.3%
1991	11.8%	13.8%	14.5%	11.0%	30.5%	5.8%
1992	7.3%	10.3%	12.3%	12.0%	7.6%	3.7%
1993	16.3%	25.4%	26.3%	23.3%	10.1%	3.2%
1994	–1.2%	–3.1%	–3.5%	–4.4%	1.3%	4.6%
1995	13.1%	7.8%	11.1%	12.5%	37.6%	6.0%
1996	13.7%	12.8%	14.4%	16.8%	23.0%	5.4%
1997	15.0%	13.7%	16.2%	17.1%	33.4%	5.7%
1998	–1.6%	–5.5%	–5.1%	1.7%	28.6%	5.6%
1999	18.9%	28.5%	26.5%	22.4%	21.0%	5.3%
2000	5.8%	2.5%	4.1%	7.3%	–9.1%	6.4%
2001	3.1%	2.8%	2.8%	5.0%	–11.9%	3.7%
2002	3.6%	1.2%	1.0%	0.6%	–22.1%	1.8%
2003	9.0%	11.4%	11.6%	10.0%	28.7%	1.2%
2004	5.8%	7.2%	6.9%	7.2%	10.9%	1.5%
2005	5.1%	7.5%	7.5%	6.4%	4.9%	3.5%
2006	9.2%	10.2%	10.4%	7.8%	15.8%	5.1%
2007	7.7%	9.7%	10.3%	9.5%	5.5%	5.2%
2008	–19.9%	–20.9%	–21.4%	–16.8%	–37.0%	2.7%
2009	9.6%	11.5%	11.5%	10.5%	26.5%	0.3%
2010	5.1%	5.5%	5.7%	4.8%	15.1%	0.3%
2011	–3.6%	–5.0%	–5.7%	0.5%	2.1%	0.2%
2012	4.2%	4.8%	4.8%	3.2%	16.0%	0.2%
2013	7.7%	9.0%	9.0%	4.6%	32.4%	0.2%
2014	3.4%	3.2%	3.4%	2.3%	13.7%	0.2%

Source: Hedge Fund Research, Bloomberg, CISDM.

32.4 FUND OF HEDGE FUNDS PORTFOLIO CONSTRUCTION

Constructing a portfolio of hedge funds as an FoF may initially appear to be a very daunting task. Questions arise, such as: What are the different strategies to which I can allocate? Are the strategies' definitions exhaustive? How much should I allocate to each strategy? How much should I allocate to each fund? Is this optimal? Such questions are forever being debated within thoughtful FoFs. This section highlights some practical issues that may be encountered when portfolios are weighted or sized, first using some overly simple allocation methods and then using some relatively advanced methods.

32.4.1 AUM-Weighted Approach

The most standard allocation process is to use an asset-weighted approach that allocates to strategies based on the AUM reported by each strategy. Unlike many standard allocation processes, however, employing AUM weights may lead to suboptimal

EXHIBIT 32.8 Sub-Period Fund of Funds Index Performance, 1990–2014

1990–1999	HFRI FOF: Conservative Index	HFRI FOF: Diversified Index	HFRI FOF: Composite Index	CISDM FOF Diversified Index	S&P 500	LIBOR
Annual return	10.7%	12.1%	13.0%	12.0%	19.0%	5.4%
Annual standard deviation	7.1%	10.7%	10.5%	8.7%	14.2%	1.4%
Excess return	5%	7%	8%	7%	13.6%	
Beta	0.08	–0.05	–0.01	0.19		
Alpha	4.3%	7.4%	7.8%	4.1%		
Sharpe ratio (R_f = LIBOR)	0.76	0.62	0.73	0.76	0.96	
2000–2014						
Annual return	3.73%	4.04%	4.11%	4.20%	6.09%	2.17%
Annual standard deviation	7.30%	8.21%	8.41%	6.62%	19.26%	2.15%
Excess return	1.56%	1.86%	1.94%	2.03%	3.92%	
Beta	0.30	0.36	0.36	0.27		
Alpha	0.37%	0.45%	0.53%	2.98%		
Sharpe ratio (R_f = LIBOR)	0.21	0.23	0.23	0.31	0.20	0.00

Source: HFR Global Hedge Fund Industry Report (2014).

allocations. If hedge funds are earning alpha (excessive risk-adjusted returns) when capital flows into a given hedge fund strategy, then under the backdrop of efficient markets, the alpha should eventually disappear. Following this reasoning, an allocation process in which the largest allocation goes to the most popular strategies, or the strategies with the most AUM, may not be optimal. In fact, some FoFs even argue that a better approach would be allocating to hedge fund strategies in which flows are shrinking or receding.

The dynamic nature of assets flowing across hedge fund strategies makes following AUM weights challenging over time. The cost of accessing this data and the fact that one vendor may monitor a hedge fund universe that vastly differs from that of another vendor or uses different strategy classifications further complicates the analysis. Therefore, even the AUM numbers are a bit suspect and may be biased to include only those hedge fund managers who may be looking to raise more assets.

Consider also that, given the vast differences in strategy definitions for hedge funds, even classifications become a bit fuzzy across time. For example, a multistrategy hedge fund with a large convertible arbitrage allocation may be initially classified as convertible arbitrage. Over time, if convertible arbitrage opportunities wane, the fund may increase allocations to equity hedge and event-driven strategies. It may not always be clear that the classification system will account for such changes.

In an example in which a hedge fund manager has initial success from his original strategy, asset inflows will follow, and the fund may branch out and add even more strategies as AUMs increase. Since hedge fund strategies are not perfectly correlated, this diversification process adds benefits on two important fronts for the manager: by lowering overall fund volatility and by increasing overall fund capacity. The largest hedge funds, therefore, often look more like multistrategy hedge funds than single-strategy funds. This typical style drift may not be properly accounted for by data vendors or may be reported on a lagged basis.

Also, if a new hedge fund strategy attracts a large amount of assets and the vendor does not have that particular strategy classified, there may be a significant amount of time before the new classification is defined. These are just some of the issues that have to be thought through carefully before pursuing an allocation process based on predefined-strategy AUM. Once strategies are defined and weights are determined, investors need to size the allocation to each of the underlying hedge funds.

Typically, FoFs are organized across strategy experts, such as a portfolio manager who covers equity hedge funds and a portfolio manager who covers global macro/commodity trading adviser (CTA) funds. Portfolio managers are generally supported by analysts who are responsible for creating a short list of funds to satisfy the targeted allocations for each strategy.

32.4.2 Equally Weighted Approach

The equally weighted approach to strategies weighting simply invests equally in each strategy. This approach to portfolio construction, as well as several others discussed in the next sections, is demonstrated in Exhibit 32.9 using empirical examples. In these examples, eight HFRX strategy indices are used to illustrate alternatives to AUM weighting strategies within an FoF portfolio.

The equally weighted approach simply gives an equal allocation to each strategy, as shown in column 1 of Exhibit 32.9. Since there are eight strategies, each strategy gets 1/8 weight, or $1/(\#\text{ of funds})$. This overly simplified example assumes both the ability to liquidate or trade any strategy each month and no transaction costs. Both assumptions are unrealistic, since it is well known that many hedge funds have lockup periods and redemption schedules and may occasionally impose gates against redemptions. The equally weighted portfolio has an annualized return of 3.9% and an annualized standard deviation of 4.8%, resulting in a Sharpe ratio of 0.81 (assuming 0% as the risk-free rate).

This approach is the most objective and simplest to implement and maintain from a rebalancing standpoint and does not require a view on any particular strategy. It does, however, treat all strategies equally and does not take into account that some strategies are more volatile than others or that some strategies have lower correlations with other strategies, which may improve diversification. Performance for this approach, which is detailed in Exhibit 32.9, is close to the average of all methods presented, but volatility is the highest. Typically, a naïve equally weighted allocation across strategies provides a good baseline on which to further improve by using one of the strategies in the next sections.

32.4.3 Equally Risk-Weighted Approach

Equally weighting across strategies does not account for the differing volatility (or annualized standard deviation) of individual strategies, and constrains a high-volatility strategy to have the same weighting as a low-volatility strategy. One way to

EXHIBIT 32.9 Possible Allocation Methods in Fund-of-Funds Portfolios

January 1998 to December 2014 (204 observations)	(1)	(2)	(3)	(4)	(5)	(6)
Strategy	**Equally Weighted**	**Equally Risk-Weighted**	**Mean-Variance Unconstrained**	**Mean-Variance Constrained 30% (Maximum)**	**Skew = 0**	**Minimum Kurtosis**
HFRX Convertible Arbitrage Index	12.5%	20.1%	0.0%	0.0%	0.0%	0.0%
HFRX Distressed Securities Index	12.5%	10.9%	0.0%	1.9%	0.0%	10.0%
HFRX Equity Hedge Index	12.5%	7.3%	0.0%	4.8%	2.4%	0.0%
HFRX Equity Market Neutral Index	12.5%	31.4%	0.0%	26.2%	30.0%	30.0%
HFRX Event Driven Index	12.5%	7.3%	0.0%	4.8%	7.6%	0.0%
HFRX Macro Index	12.5%	6.5%	11.3%	21.3%	30.0%	30.0%
HFRX Merger Arbitrage Index	12.5%	7.0%	88.7%	30.0%	30.0%	30.0%
HFRX Relative Value Arbitrage Index	12.5%	9.5%	0.0%	11.0%	0.0%	0.0%
Annual return	3.9%	2.9%	5.3%	4.0%	4.1%	3.9%
Annual standard deviation	4.8%	4.6%	3.4%	3.4%	3.6%	3.5%
Skew	–2.8	–3.8	–0.9	–0.7	0.0	–0.1
Kurtosis	15.6	25.6	2.7	2.3	1.1	1.1
Sharpe ratio ($R_f = 0\%$)	0.81	0.64	1.56	1.20	1.15	1.14

Source: CISDM and authors' calculations.

normalize for the different volatilities would be to size the allocations inversely proportional to each strategy's volatility. For example, highly volatile strategies would get a smaller allocation, while less volatile strategies would get a larger allocation. In practice, some analysts employ rolling standard deviations, but then the choice of window length and frequency of data comes into question: Should one use the data at daily, monthly, or annual intervals, and how many observations should be used? When hedge funds have experienced several months of extremely poor performance, including or excluding one monthly data point will have a profound influence on the rolling standard deviations used in certain times. Another practical issue is the periodic reallocation across strategies, which may take some time given underlying fund liquidity terms.

With the many different ways used to compute volatility, this process has many options. Standard deviation is an imperfect measure of volatility because it penalizes upside deviations as much as downside deviations. One of several alternatives to standard deviation when estimating and sizing allocations is to use semideviations, which measure only downside volatility.

Equally risk-weighted allocations are shown in column 2 of Exhibit 32.9. These allocations are constructed by taking the inverse of each strategy's whole-period annualized standard deviation. The sum of the inverse standard deviations is the denominator, and the strategy's inverse standard deviation is the numerator. Thus, the weights sum to 1.0 and can be directly compared to the equally weighted allocations:

$$\text{Equally Risk Weighted}_i = \frac{(\text{Standard Dev of Fund } i)^{-1}}{\sum_{K=1}^{N} (\text{Standard Dev of Fund } k)^{-1}} \tag{32.1}$$

This approach is a simplified version of the risk parity approach. In a risk parity approach, each asset class makes the same contribution to the total volatility of the portfolio. Assuming that all asset classes, or in this case all hedge funds, have the same correlation with each other, then the weights obtained in Equation 32.1 will be the same as risk parity. That is, the weights are selected so that each fund will make the same contribution to the total volatility of the portfolio. Once the inputs to the model have been determined and a rebalancing discipline has been created, this allocation process is easy to maintain and provides some comfort to adherents since the equally risk-weighted approach incorporates elements of risk. Its biggest drawbacks are that it uses historical information to allocate capital without considering whether markets have changed, doesn't take into account any cross-correlations with other strategies, is very dependent on the time frame selected, and, in times of particular stress in the markets, produces results that may not be representative of long-term volatility levels.

In this example, the HFRX Equity Market Neutral Index has a substantially greater allocation (31.4%) than the HFRX Macro Index, which has the lowest risk weight (6.5%), given its higher relative volatility. The equally risk-weighted portfolio has an annualized return of 2.9% (lowest), an annualized standard deviation of 4.6%, and a Sharpe ratio of 0.64 (lowest). In addition, it has the most unfavorable skew and kurtosis of the four methods shown.

32.4.4 Mean-Variance (Unconstrained and Constrained) Approach

Probably the best-known allocation method is based on mean-variance optimization (MVO), as shown in columns 3 and 4 of Exhibit 32.9. This method involves finding optimal allocations across strategies that increase the ex post Sharpe ratio. Some have argued that MVO, which uses historical data for inputs, is error maximizing, as it overly favors historically strong-performing investments and overly emphasizes prior diversification benefits. In addition, MVO assumes that returns are normally distributed, which is not the case for hedge funds. These are legitimate reasons to be wary of the MVO allocation approach, which can create optimal, but not practical or feasible, portfolios.

Constraints are often used to improve diversification by reducing the dominant weights that unconstrained optimization may suggest for some strategies. Results such as those shown in column 4 of Exhibit 32.9, with a 30% maximum constraint or some other reasonable constraint, are common. Using HFRX index data, the unconstrained optimization in column 3 yields an oversized 88.7% allocation to the HFRX Merger Arbitrage Index and an 11.3% allocation to the HFRX Macro Index.

The MVO unconstrained portfolio has an annualized return of 5.3% (highest), an annualized standard deviation of 3.4% (tied for lowest), and a Sharpe ratio of 1.56 (highest). The MVO constrained portfolio, which is widely used in practice as opposed to the unconstrained model, has a lower annualized return of 4.0%, the same annualized volatility of 3.4%, and a slightly lower Sharpe ratio of 1.2. The skew and kurtosis of both MVO approaches are roughly the same. While the results of the MVO unconstrained model appear to be attractive on the surface, the lack of diversification among strategies subjects the portfolio to tail risk events.

32.4.5 Mean-Variance with Constraints on Higher Moments Approach

Mean-variance optimization may not be appropriate, as the returns for hedge fund strategies are not normally distributed, as is evidenced by negative skewness and excess kurtosis. Black (2006, 2012) documents skew-reduction techniques by including the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) in a hedge fund portfolio. In columns 5 and 6 of Exhibit 32.9, we optimize the allocations by forcing skew and excess kurtosis close to zero. Interestingly enough, the underlying strategies have sufficient distributional characteristics that allow portfolio skewness and kurtosis to be substantially improved relative to the portfolios designed by the equally weighted or MVO approaches.

Note that the portfolios in columns 4 and 5 assign maximum weights to merger arbitrage, global macro, and market neutral strategies, while avoiding allocations to convertible and fixed-income arbitrage strategies that are known to have unfavorable characteristics (i.e., undesirable higher moments). Limiting the maximum weight on a strategy, such as is done in column 5 in Exhibit 32.9, will bring more diversification to the low skewness/excess kurtosis allocations.

There are many objectives that can be satisfied using portfolio optimization, including minimizing standard deviation, minimizing drawdown, or minimizing the correlation or beta to specific market indices. Optimization methodologies assume the stability of the return distribution over time, which tends to break down during fat tail events, especially across hedge fund strategies. As a result, much caution should be taken when using these methodologies. Some analysts will use volatility and correlation estimates from times of market crisis to understand the sensitivity of optimization results to inputs that can vary substantially over time.

It is important to note that the allocations in Exhibit 32.9, columns 3 through 6, have been made ex post, so the returns could not have been realized. These portfolios were created with hindsight, meaning the results are not realistic or achievable by any investor. This exercise was purely for the purpose of offering insight into alternative portfolio allocation processes typically employed in the FoF industry. In the end, such outputs would typically be generated and a group discussion would ensue within the FoF's investment and risk committees. Ranges are often used, rather than hard numbers, and allocations may depend on available (open) funds that have been preapproved for investment. In addition, this analysis should be performed ex ante to take into account the underlying hedge funds' liquidity terms.

32.4.6 Personal Allocation Biases Approach

When an FoF manager has had a very bad experience with a given strategy, typically due to performance or operational issues, the manager's personal biases will have a direct influence on future asset allocation decisions. Some FoFs avoid mortgage-backed and fixed-income arbitrage funds altogether, while others may exclude all quantitative or black-box models. Institutional investors may have written mandates to avoid certain strategies that may have blown up in the past. Such biases are conspicuously present across asset allocation methodologies in practice, as legacy experiences weigh in on current decisions.

32.5 MANAGER SELECTION

There are many ways to populate each strategy allocation with hedge funds. Typically, the portfolio manager or strategy expert searches for the combination of hedge funds that can best meet the objectives of the FoF. This process, however, involves many factors, including whether the identified hedge fund is open to new investment. Unlike mutual funds, which are easy to access for investors, hedge funds are not always open to new investors. Some of the best hedge funds are closed to new investors after reaching the manager's capacity constraint, the AUM level beyond which the manager believes that additional capital will increase market impact and reduce potential returns. Also, the returns to some strategies can be closely tracked with just a few funds, while other strategies may require a larger number of funds.

FoF portfolio managers are constantly looking to improve their current hedge fund rosters by building a short list of hedge funds approved for investment. Occasionally, an underlying fund needs to be replaced, which entails coordinating redemption cash flow timing as well as reallocating capital. FoF managers are constantly attending conferences, capital introduction events, networking events, and endless meetings with managers, all in an effort to keep their short lists updated with quality managers.

FoFs will also subscribe with data vendors and may set up quantitative screens to help make their searches more efficient. An example of such a methodology is the Bifurcated Fund Analysis Model (BFAM) by Liew, Mainolfi, and Rubino (2002). This model helps determine whether a potential hedge fund manager's return characteristics are a good fit from two vantage points: (1) the attractiveness from a mean-variance perspective, and (2) the rankings from a flexible peer group scoring methodology.

32.6 BENEFITS OFFERED BY FUNDS OF HEDGE FUNDS

As already mentioned, funds of hedge funds charge additional fees to investors who wish to use them to invest in hedge funds. Therefore, it is natural to ask why investors choose to use a fund of funds platform to allocate to hedge funds. As described next, some investors may lack the expertise, resources, access, and size to use the direct allocation approach. The fees earned by funds of funds represent compensation for the benefits that they provide to investors. Twelve benefits are:

1. EXPERTISE: FoF managers have experience in understanding both complex strategies and the universe of managers that execute these strategies. In addition, hedge fund managers may run several variations of specific funds, which adds further complexity to an already information-constrained undertaking. Because of their role as asset allocators, FoFs have the ability to access, collect, and interpret data gleaned from various channels, including data providers, prime brokers, and industry contacts. This, plus constant dialogue with hedge fund managers, gives them an informational advantage over nonprofessional investors.
2. MONITORING: In their study of the value of registered FoFs, Aiken, Clifford, and Ellis (2014) find only limited evidence that FoFs exhibit skill when selecting hedge funds. Conversely, they find strong evidence that FoFs make skillful termination decisions. They also find an informational asymmetry in favor of incumbent investors versus prospective investors, due primarily to the ongoing learning about a specific hedge fund through the monitoring process.
3. IMMEDIATE EXPOSURE: FoFs offer investors an easy option for getting the desired immediate exposure to a group of hedge fund strategies that is managed by experienced professionals.
4. DIVERSIFICATION: Prudent investing dictates that portfolios should be well diversified. Some investors lack the necessary capital and/or expertise to invest directly in hedge funds, which would allow for appropriate diversification and risk reduction. The diversification level of an FoF portfolio, however, is not necessarily a function of the number of underlying strategies or funds. This is because hedge funds are not single securities but represent a diversified portfolio of securities. A key to understanding the degree to which an FoF portfolio is truly diversified is to analyze each underlying hedge fund to determine if any meaningful factor exposures are held in common with the other funds in the portfolio.

   Numerous researchers have studied the impact of increasing the number of funds in an FoF portfolio on various risk measures using a naïve diversification approach (i.e., random selection and equal weighting of the underlying managers). For a single-strategy FoF, an equally weighted portfolio consisting of approximately three to five hedge funds is usually highly correlated with its respective hedge fund strategy index. This clearly supports the idea, say researchers, of running relatively concentrated single-strategy FoFs. The same has been found to be true for multistrategy FoFs. Lhabitant and Learned (2003, 2004) suggest that an equally weighted portfolio of only 10 to 15 hedge funds is sufficient to diversify away most of the underlying manager-specific risk. Researchers conclude that this also supports the strategy of running concentrated multistrategy FoFs.

5. Accessibility and economies of scale: The price of entry for investing in a single hedge fund is often $1 million, which makes doing so unaffordable for most individual investors. By comparison, minimum investment levels for FoFs are relatively low. This allows more individual investors and small institutions to gain access to hedge funds, even though their investment capital may be comparatively small. Moreover, investors essentially share the costs associated with managing an FoF with their co-investors.
6. Liquidity: Investments in hedge funds are relatively illiquid, due to initial lock-ups, limited redemption dates, notice periods, and potential redemption restrictions available to managers. By comparison, the liquidity terms offered by FoFs are seen by some as more compelling. Most FoFs offer quarterly or monthly liquidity under normal market conditions. Some FoFs even offer daily liquidity, either through a listing on an exchange or via an over-the-counter secondary market that matches supply with demand.
7. Access to specific managers: Access to the best talent and ideas in the hedge fund community is a scarce resource, and often the most desirable hedge funds are closed to new investments. Most investors do not have the necessary networks and/or relationships to get investment capacity in these funds when it does become available. Investing in an existing FoF that is already allocated to these desirable hedge funds is the fastest way to immediately participate in their performance.
8. Negotiated terms: Thanks to the power of their collective assets, some FoFs have successfully negotiated access to certain hedge funds at reduced fees with improved liquidity and/or transparency terms. This is normally beyond the capabilities of most individual investors.
9. Knowledge transfer: Many first-time institutional hedge fund investors look at FoFs not only as an investment vehicle but also as a way of learning about hedge fund strategies and managers. These investors typically switch to direct investments in hedge funds once they have the experience to go direct.
10. Regulation: In order to facilitate distribution to a wider audience, some FoFs choose to register in regulatory jurisdictions that offer better investor protection than their underlying investments, even though the cost and administrative and operational burdens may be higher. This can often be reassuring for first-time investors and helps promote improved transparency, oversight, and reporting.
11. Currency hedging: While the currency of choice in the hedge fund world is the U.S. dollar, many FoFs offer share classes denominated in various currencies by hedging the currency risk. Although institutional investors often wish to manage their own currency risks, many small or private investors who prefer to be shielded from currency fluctuations can invest in share classes denominated in other currencies when available.

12. Leverage: Some FoFs provide leverage to their investors. In addition to the capital provided by their investors, FoFs may borrow money to invest in the portfolio. This allows them to produce higher returns in low-volatility periods than would an unlevered FoF, as long as the interest costs are lower than the unlevered returns of the underlying hedge fund portfolio.

32.7 DISADVANTAGES OF FUNDS OF HEDGE FUNDS

The potential benefits of funds of funds were described in the preceding section. Here, we list and discuss seven potential disadvantages of using funds of hedge funds to allocate to hedge funds. Of course, having to pay extra fees is the most important disadvantage, but, as we see, there are others as well.

1. Double layer of fees: The **double layer of fees** in FoFs refers to the two layers of fees charged to investors, including those at the fund level and at the fund of funds level. FoF managers effectively pass on to their investors all fees charged by the underlying hedge funds, while also charging a set of fees for their services and those of service providers to the FoF. As an illustration, Ang, Rhodes-Kropf, and Zhao (2008) estimate that in the TASS database, the average FoF charges a 1.5% management fee and a 9.2% performance fee on top of the underlying hedge fund average management and incentive fees of 1.4% and 18.4%, respectively. However, due to the competition from capable consultants, coupled with the poor performance of FoFs in 2008, institutional investors making large allocations have recently been paying much lower fees to FoF managers than the historical 1% management fee and 10% incentive fee; in fact, FoFs that charge incentive fees are no longer competitive for institutional mandates.
2. Lack of control: In a typical FoF, investors give up control over how the assets are invested, and because they may not have a direct relationship with the hedge funds in which the FoF invests, their transparency is impaired.
3. Lack of transparency: Some FoF managers do not disclose the holdings of their portfolio or their strategy allocation. They contend that these represent their intellectual property, which they are reluctant to share with investors, as the information may eventually end up in competitors' hands to the detriment of their own performance. In such cases, it becomes relatively difficult for FoF investors to understand how the performance is being generated and how risk is changing in the portfolio. Institutional-quality FoFs offer more transparency to their investors.
4. Lack of customization: Smaller or less sophisticated investors who rely on FoFs to access hedge funds typically invest in a portfolio of hedge funds with a broadly stated mandate. Larger investors, who have the critical mass and sophistication, can work with FoF managers to create their own fund-of-one portfolio, which is customized to meet their more precise objectives.
5. Netting risk: In an FoF, the investor must pay incentive fees to each of the profitable underlying hedge funds, regardless of the performance of the overall portfolio. **Netting risk** is an investor's exposure to economic loss from paying higher total fees as a result of paying incentive fees for funds with profits without offset reductions from funds with losses. Thus, netting risk is incurred when, for

example, half the funds are down 10% and the other half are up 10% on a gross basis, and the FoF will still have to pay an incentive fee to the positive-performing hedge funds despite no return at the aggregate level.

6. Co-investor risk: FoFs commingle the assets of a number of investors who may have different investment objectives and time horizons. As a result, **co-investor risk** occurs when co-investors are jointly affected by inflows and outflows, the latter of which may trigger an increase in leverage to finance redemptions when cash or liquidity in the underlying funds isn't readily available. Another potential solution to satisfy investors' requests for redemptions would be for the FoF manager to sell the most liquid funds first, leading to a potential change in the FoF's stated objectives. **Funds of one**, or custom portfolios for a single investor, are typically not exposed to these issues.
7. Administrative delay risk: FoFs need time to get information from each of the underlying hedge funds and aggregate it for reporting purposes, which means investors get FoF information with a lag. This is especially challenging around year-end, when each of the underlying funds is audited and tax returns (where applicable) need to be prepared. **Administrative delay risk** is economic exposure to consequences from tardy reports and other information. In some cases, the time lag is of sufficient magnitude that investors may be required to file extensions for U.S. tax reporting purposes.

32.8 FUNDS OF HEDGE FUNDS VERSUS MULTISTRATEGY FUNDS

An FoF is a pooled investment vehicle that aggregates investor capital in a fund that invests in a diversified group of hedge funds, run by external managers who employ diverse strategies to meet the stated objectives of the FoF. While the FoF is subject to the terms of each individual fund into which it invests, some investors will request a side letter, which details how the terms of the underlying hedge fund for a specific investor will differ from the terms of other investors. Investors requesting side letters often ask for specific fees, differential liquidity terms, or a focus or prohibition on specific areas of investment.

The FoF manager charges a fee, in addition to those of the underlying managers, for managing the FoF. Whenever possible, FoFs will negotiate with the underlying hedge funds for a reduced fee, typically in return for a sizable allocation or a longer lockup period on the investment. Negotiated fee reductions flow through directly to investors in the form of lower overall fees paid and higher net-of-fee performance. Generally, the FoF does not use leverage at the fund level, but some underlying hedge funds may use leverage as part of their strategy.

Initially, during the 1990s, FoFs were attractive because they were experts in the mysterious world of hedge funds and produced attractive returns that were uncorrelated to the markets. In the 2000s, FoFs were used by institutional investors to gain exposure to hedge fund strategies and to learn about hedge funds (knowledge transfer). Finally, FoFs were used to shield institutional investors from negative headlines when an underlying hedge fund blew up. Today, the hedge fund industry is more transparent and institutional investors are so prevalent that few mysteries remain; thus, the value proposition of FoFs for institutional investors has diminished over time.

Multistrategy funds (MSFs) are a type of hedge fund that invests its capital across several strategies that are run by different internal portfolio managers, or traders, at the hedge fund. MSFs identify talented portfolio managers that are complementary to the other managers at the MSF, and give them a platform that allows them to focus exclusively on managing money. In exchange for managing a portion of the MSF's assets, the MSF traders will keep a prenegotiated portion of the profits they generate. The MSF manager is responsible for allocating capital internally to maximize risk-adjusted returns, for providing a robust infrastructure that can handle disparate strategies, and for managing risk across the entire portfolio. Risk management is at the core of the MSF, since each strategy has the potential to take the fund down. The two most prominent examples of risk management failures are Long-Term Capital Management (1998) and Amaranth Advisors, LLC (2006).

The MSF manager charges a management fee plus an incentive fee that is based on the aggregate net performance of the underlying strategies. As opposed to an FoF, an MSF typically charges incentive fees on its net gains. In many cases, managers will also pass certain expenses of the fund on to investors, which reduces returns. These costs are passed on to investors because MSFs demand large investments in technology and infrastructure that can handle a multitude of trading strategies. In addition, many of the strategies traded focus on arbitrage, which tends to employ leverage. Risk management systems need to be robust enough to constantly provide a picture of individual traders' portfolios and composite fund risk exposures in order to enable hedging by the MSF manager when necessary. Finally, depending on the compensation model used by the MSF, profit payouts to individual teams may be included in expenses. These structural and compensation expenses can be meaningful, especially in a low-return environment.

Many hedge funds began as single-strategy funds, which used the expertise developed by their founders to give them an edge in the specific area in which they traded. For example, managers with experience in merger arbitrage might have set up a hedge fund to monetize their expertise at producing attractive returns. Over time, many strategies (typically those with the word arbitrage in their name) found that there would be long periods of time when the specific strategy would provide little or no returns. As a result, astute managers sought to create MSFs, with which they would be able to generate returns in most market conditions by tactically rotating allocations across strategies. This format also allowed them to accommodate larger amounts of capital by increasing the number of investment teams on the platform.

There are four key issues that need further analysis when deciding between FoFs and MSFs:

1. FEES: FoFs charge an additional layer of fees in addition to the fees charged by the underlying hedge funds. In many cases, they are able to negotiate fee breaks with underlying managers as a way to offset the additional layer of fees at the FoF level. Many MSFs, which do not have the additional layer of management and incentive fees, may pass on the cost of their infrastructure to the fund, and in many cases, these costs are substantial. However, MSF fees may be lower

than those of FoFs and the underlying managers, if for no other reason than the reduction of netting risk.

2. Leverage: While both FoFs and MSFs may use leverage, it's important to understand how that leverage is used, its tenor (how long it is locked in), and its impact on the overall fund. Leverage in an FoF can cause losses when an individual fund blows up, but those losses are usually limited to specific underlying funds and do not contaminate the rest of the portfolio. Leverage in an MSF, in contrast, can put the entire fund at risk, as the strategies are not isolated unless housed in a series of bankruptcy-remote special purpose vehicles.
3. Tactical moves: FoFs typically take longer to make tactical moves, as they must provide notice to redeem, wait for the funds, and then reallocate to new funds. Since the capital in an MSF is internal, MSFs can make tactical moves more quickly to take advantage of market conditions, if a team is already in place to execute the strategy.
4. Compensation: While MSFs' compensation arrangements with their trading teams vary widely across firms, they generally fall into one of two categories. In the first category, MSFs pay their internal teams based on their aggregate performance. This was the original compensation model for MSFs, but it created a talent-retention issue when the best traders decided to strike out on their own. In the second category, MSFs pay each internal team based on the profits each generates. This approach is the most widely used compensation structure today because it helps retain top talent and is aligned with investors' interest in paying for performance. Unfortunately, it also creates netting risk for investors when profitable teams in an MSF get paid even though the aggregate fund reports negative returns. These compensation charges are frequently included in expenses that are passed on to investors.

Exhibit 32.10 highlights major differences between FoFs and MSFs.

32.9 HOW FUNDS OF HEDGE FUNDS ADD VALUE

One of the most important debates with respect to FoFs concerns whether they deserve their "fees on top of fees" and whether they add any value with respect to a randomly selected portfolio of, say, 20 to 40 hedge funds. In practice, there are essentially three ways for an FoF manager to add value.

32.9.1 Adding Value through Strategic Allocation

The strategic allocation approach takes a longer-term view to allocating to various hedge fund strategies without varying the weights over time. Running a FoF is not simply a matter of assembling a large collection of good managers, as that can still result in a concentration of risks if there is a high level of correlation in the underlying factor exposures of these funds. Building a top-down strategic allocation typically involves analyzing the expected long-term risk-return profiles of the different strategies, as well as estimating the correlation of their expected returns. The goal is to create an initial portfolio allocation consistent with the fund's long-term objectives

EXHIBIT 32.10 Funds of Hedge Funds Compared with Multistrategy Funds

	Funds of Hedge Funds	Multistrategy Funds
Asset allocation	Open architecture; can select from hedge fund universe	Captive managers
Portfolio diversification	More diverse in strategies, styles, and funds (for diversified FoF)	Narrower in strategies and styles
Investor sophistication level	Low	High
Minimum capital investment	Low	High
Headline risk	Lower, but market generally aware of underlying manager allocations	Higher; each captive manager puts fund at risk
Transparency	Typically, fund names, strategies, and allocation percentages disclosed	Varies, but typically strategy allocations and weights disclosed
Performance	Lower	Higher
Returns volatility	Lower	Higher
Operational risk	Low	High
Cross-collateralization (at what level leverage is applied)	Underlying fund level	At the strategy and/or overall fund level
Basis of risk analysis	Generally at the underlying hedge fund level	Security level
Ability to make tactical changes	Takes longer to redeem and reallocate	Capital can be moved quickly to take advantage of market conditions
Risk management	Important	Critical
Fund liquidity	Moderate	Moderate
Standard fees:		
Management fee (plus underlying HF fees)	1.0%	2.0%
Incentive fee	Up to 10%	20%
Pass-through expenses	No	Varies, but yes in many cases
Are incentive fees paid on net returns?	No	Yes, mostly
Knowledge transfer to investors	Yes	No

Source: CISDM.

and constraints. This task determines the long-run beta of the fund with respect to various sources of risk.

32.9.2 Adding Value through Three Key Decisions in Tactical Allocation

Tactical allocation refers to a shorter-horizon active strategy that seeks to enhance portfolio performance by opportunistically shifting the asset allocation in response to changing market dynamics. Many FoFs follow a top-down tactical allocation process, which frequently involves making three key decisions:

1. Strategy weightings: Overweighting or underweighting specific investment strategies
2. Timing: Deciding when the changes should be made (for example, based on certain indicators or factors)
3. Sizing: Determining the appropriate size of the contemplated changes

In practice, however, an FoF is limited because of the liquidity terms of the underlying hedge funds unless it invests in only the most liquid areas of alternative investments, uses liquid alternative mutual funds, or uses separately managed accounts (SMAs). Fund terms typically determine how long it takes for an FoF to adjust its exposures in order to take advantage of changes in the market environment.

32.9.3 Adding Value through Fund Selection

The fund selection approach focuses primarily on selecting individual funds and sizing each allocation to produce multiple streams of uncorrelated returns. While this seems very similar to a traditional stock selection activity, the reality is that FoF managers often have to make a trade-off between their ability to add value through dynamic manager allocations in highly liquid funds and the higher potential contribution to performance of less liquid funds.

32.9.4 Evidence Regarding Value Added by Fund-of-Funds Managers

A recent study by the EDHEC-Risk Institute (Darolles and Vassie 2010) sought to quantify the impact of these three factors over time (see Exhibit 32.11).

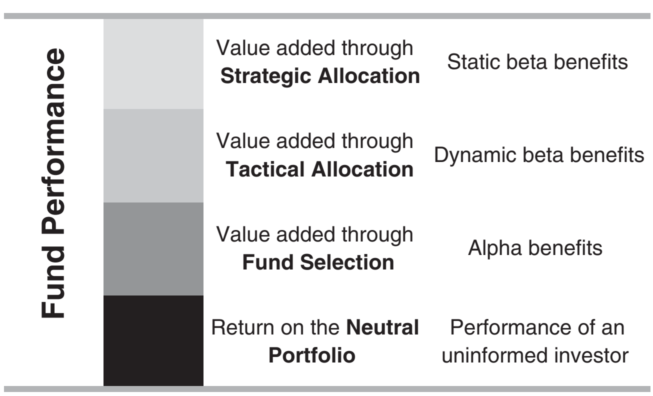


EXHIBIT 32.11 Approaches to Add Value by Fund-of-Funds Managers
Source: Darolles and Vassie (2010).

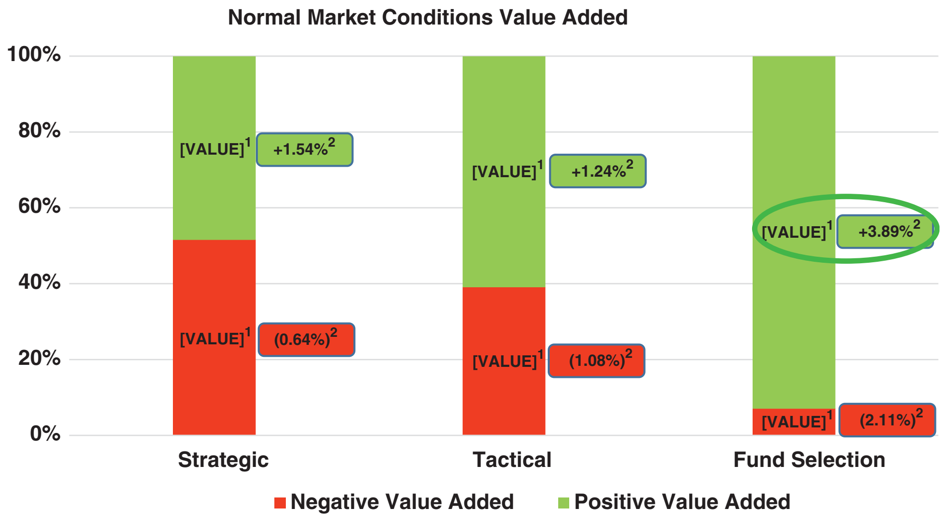


EXHIBIT 32.12 Three Approaches to Add Value by FoF Managers in Normal Market Conditions

[1] Percentage of FoFs in study.

[2] Value added over (detracted from) neutral portfolio.

Source: Darolles and Vassie (2010).

From a universe of more than 1,000 FoFs, Darolles and Vassie (2010) studied the performance attribution of 184 FoFs that had a continuous track record from January 2000 to July 2009. They compared the results of these FoFs against the returns of a predefined neutral portfolio of an uninformed investor (as defined in the study). The research concluded that of the three approaches, strategic allocation accounts for a large part (68%) of FoF return variability, with the balance explained by the other two factors.

The second part of the study sought to assess the extent to which the sources of value added over the neutral portfolio by FoF managers are regime dependent. The study divided the same observation period into two market regimes: normal market conditions (January 2000 to June 2007) and stressed market conditions (June 2007 to July 2009). These are displayed in Exhibit 32.12 and Exhibit 32.13. In addition, the study classified FoF performance into one of two groups depending on their performance versus the neutral portfolio: positive value added and negative value added. In other words, which of these three approaches added the most value during normal market conditions and during stressed market conditions?

The researchers made the following conclusions:

1. Strategic allocation value added is positively skewed in both market regimes. In normal market conditions, 48.4% of FoFs added 1.54% annually over the neutral portfolio by emphasizing strategic allocation, but it was in stressed market conditions that this approach was most valuable, as 77.7% of FoFs were able to add 3.50% annually. In both regimes, FoFs classified as negative value added hurt performance by less than 1.0%. It is precisely when it is needed the most that benefits from strategic allocation are strongest, which is not surprising given that strategic allocation and risk management are interrelated.

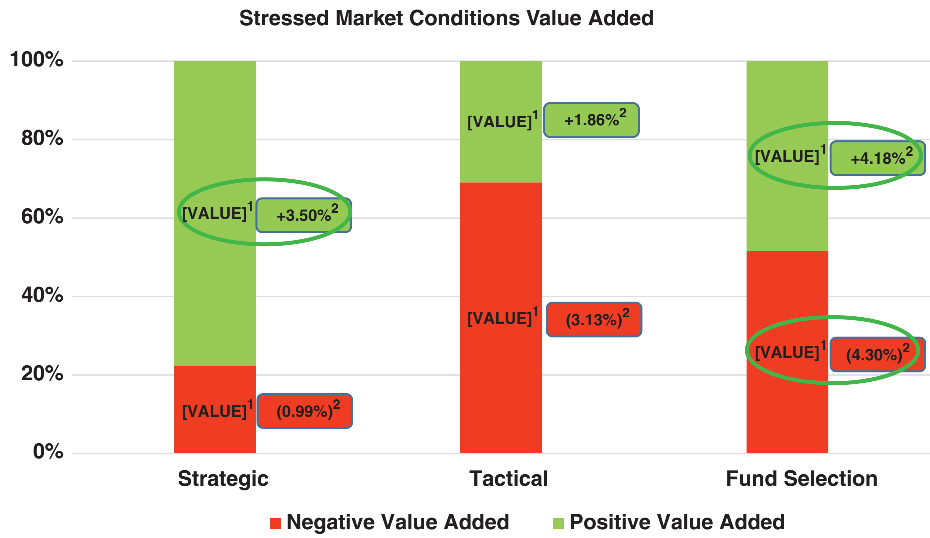


EXHIBIT 32.13 Three Approaches to Add Value by FoF Managers in Stressed Market Conditions
[1] Percentage of FoFs in study.
[2] Value added over (detracted from) neutral portfolio.
Source: Darolles and Vassie (2010).

2. Tactical allocation appears to produce limited value added. By tactically adjusting the portfolio in normal market conditions, 60.9% of FoFs added 1.24% annually over the neutral portfolio, costing the other 39.1% of FoFs 1.08%. Conversely, in stressed market conditions, only 30.9% of FoFs were able to add an average of 1.86% annually, while it cost the other 69.1% of funds 3.13% annually to use the tactical allocation approach.
3. Fund selection is a double-edged sword. In normal market conditions, 92.9% of FoFs had an annual positive outperformance of 3.89% over the neutral portfolio by using the fund selection approach. In stressed market conditions, however, only 48.4% of FoFs added value of 4.18% annually, while the other 51.6% cost their FoFs 4.30% annually. Fund selection exhibits an unfavorable regime dependency, as it is precisely when it is counted on to boost performance that it may add to downside risk.

To analyze whether actively managed FoFs have, on average, generated substantial value added over noninvestable hedge fund indices, the performance of the CISDM Fund of Funds Diversified Index can be compared to the performance of the CISDM Equal Weighted Hedge Fund Index (see Exhibit 32.14).

FoFs provide significant diversification potential because they have relatively low volatility and drawdowns. This is particularly true of FoFs as compared to individual hedge funds. This suggests that a fiduciary who is primarily concerned about the downside risk associated with hedge fund investments should consider an FoF vehicle. This risk reduction, however, generally comes at the cost of lower annualized returns, because average returns of FoF indices have typically been only a little more

EXHIBIT 32.14 Analyzing Fund of Funds Performance versus Hedge Fund Strategy and Other Market Indices

January 1990 to December 2014 (300 observations)	Annualized Mean	Annualized Standard Deviation	Skewness	Information Ratio	Maximum Drawdown	Correlation to CISDM Fund of Funds Diversified Index
S&P 500 Total Return Index	10.3%	14.6%	–62.5%	70.5%	–50.9%	55.0%
MSCI ACWI Index	5.6%	15.4%	–68.3%	36.3%	–56.2%	59.8%
MSCI Emerging Markets Index	8.8%	23.2%	–68.6%	37.8%	–62.7%	65.2%
J.P. Morgan Global Bond Aggregate Index	6.4%	5.8%	12.8%	111.1%	–9.4%	7.2%
MSCI EAFE U.S. Currency TR Index	3.6%	17.3%	–46.4%	21.0%	–58.2%	54.3%
CISDM Fund of Funds Diversified Index	6.9%	4.8%	–71.9%	144.7%	–17.5%	100.0%
CISDM Equal Weighted Hedge Fund Index	11.2%	7.1%	–55.2%	158.4%	–21.7%	87.6%
CISDM Merger Arbitrage Index	8.4%	3.8%	–90.5%	222.1%	–5.7%	65.8%
CISDM Global Macro Index	9.0%	5.5%	136.3%	164.5%	–8.2%	60.6%
CISDM Event Driven Multi-Strategy Index	10.3%	5.7%	–149.2%	179.8%	–20.2%	80.0%
CISDM Equity Market Neutral Index	7.6%	2.0%	–40.8%	378.5%	–2.8%	62.2%
CISDM Equity Long/Short Index	10.7%	7.8%	–27.6%	137.0%	–17.2%	78.6%
CISDM Distressed Securities Index	11.1%	6.6%	–137.7%	168.0%	–21.2%	66.2%
CISDM Convertible Arbitrage Index*	8.8%	4.7%	–358.9%	187.6%	–22.5%	61.8%

*Data from January 1992 to December 2014.
Source: CISDM and Bloomberg.

than half of those of hedge fund indices over the same period. This is in line with the empirical literature and can be explained by the so-called double layer of fees of FoFs and the survivorship impact that creates a meaningful upward bias in the reported performance of individual hedge fund indices.

On a risk-adjusted basis, research also indicates that FoFs offer a slightly lower information ratio. This suggests either that FoF managers have not done a particularly good job at selecting superior hedge funds or that the fees they charge wipe out the benefits they deliver. Several studies have confirmed these results and tend to confirm that, on average, FoFs underperform hedge funds after fees. See Ackermann, McEnally, and Ravenscraft (1999); Kat and Lu (2002); Brown, Goetzmann, and Liang (2004); Capocci and Hubner (2004); and Fung and Hsieh (2004).

These results should be taken with caution, however. Comparing average risk-adjusted returns or alphas across two asset classes is valid only if they are directly comparable. Single-manager hedge funds and FoFs are not, however, directly comparable. As stated by Ang, Rhodes-Kropf, and Zhao (2008), the average hedge fund is not an investable entity because the best hedge funds are often closed to new investors. Even if an investor meets the high minimum requirements for investing in hedge funds, there is no guarantee that a successful hedge fund will take that investor as a client.

If there were no FoFs, all unskilled investors would be forced to invest directly in hedge funds, including the bad ones, which would not receive any funding in a world where FoFs existed. As a result, most hedge funds receive funding either from skilled investors or from skilled FoFs. **Funding bias** is noted when the returns to hedge funds observed in databases are biased upward compared to the full hedge fund universe that would have existed if there were no FoFs. That is, by their mere presence, FoFs discourage bad hedge funds from being formed or, if formed, from staying in business too long. Funding bias provides an important monitoring and due diligence function for the industry. The funding bias in hedge funds is very different from reporting biases, which are based on whether funded hedge funds report or do not report to a database. By contrast, the funding bias involves the unobserved unfunded set of hedge funds.

According to Ang, Rhodes-Kropf, and Zhao (2008), a true FoF benchmark should include both funded and unfunded hedge funds available to unskilled investors. In addition, investors experience significant costs when accessing hedge funds, including sourcing and due diligence costs, which are hard to measure but should nevertheless be estimated and factored in. When these elements are taken into account, Ang, Rhodes-Kropf, and Zhao conclude that FoFs, on average, deserve their layer of fees; that the more skilled investors are, the less likely they will be to find FoFs valuable; and that the less risk averse an investor is, the less value an FoF will provide. The conclusion is that FoFs have a **constituency effect**, in which investors with lower wealth, less skill, and greater risk aversion are natural clients for FoFs. Institutions with larger assets, greater skill, and less risk aversion are more likely, in the long run, to build a portfolio of direct hedge fund investments.

32.10 HEDGE FUNDS INDICES

In the universe of traditional assets, indexing has long been an ideal method of achieving broad-based, low-cost passive exposure to global equity and bond markets. As a result, there are countless investment vehicles based on the idea of economically tracking broad market indices, many of them with large AUM. Until recently, indexing has not been applied to the world of hedge funds, since hedge funds were historically marketed as following predominantly alpha-oriented strategies that exploit idiosyncratic risk to generate absolute returns. The increased demand for hedge fund strategies from institutional investors has progressively led to innovations that focus less on alpha and more on hedge fund beta, through identifying and capturing hedge fund beta factor exposures using indexing and passive strategies. These passive strategies offer superior transparency, improved liquidity, and lower fees.

Liew (2003) argues the case against hedge fund index investing. First, he shows that aggregate hedge fund index exposure is unwarranted, since only one-third of hedge funds have statistically significant skills when employing a single-factor, summed beta capital asset pricing model (CAPM). Next, he presents evidence to show that diversification benefits of hedge fund indices disappear in extreme market conditions. Finally, simulations reveal that with even a reasonable degree of discernment, a portfolio of hedge funds can be constructed to outperform most hedge fund indices.

32.10.1 Desirable Characteristics of Indices

Before discussing hedge fund indices in detail, let us recall the desirable properties of investment indices in general. Hedge fund indices that are used as yardsticks for investments should have the following six characteristics:

1. Representativeness: The index should accurately reflect either the whole universe of hedge funds or a universe focused on a particular style as mandated by the index.
2. Unambiguousness: The specific hedge funds included in the index must be identified by name, and their weight in the index, strategy classification, and other relevant information must be provided.
3. Verifiability: Users should be able to verify all information and calculations used to construct the index. This is greatly facilitated when the construction methodology is publicly available.
4. Accountability: The construction and the revisions of the index follow exact guidelines, which should be approved by an independent committee.
5. Replicability: Investors should be able to replicate the index with reasonable costs and minimal tracking error. Note that investability may be a subjective criterion, as it may require large amounts of capital to replicate the index.
6. Reasonableness: The index should contain only funds that are reasonable investments for a typical client and that are consistent with a prudent investor's risk preference. The weighting and rebalancing rules should be compliant with the terms of the underlying investment funds.

Although it is not difficult to create representative, investable indices of traditional assets such as stocks and bonds, this goal is not attainable in the hedge fund world. Investments in hedge funds are lumpy, which means a truly diversified and representative investable index would require an enormous amount of capital to fund as well as resources to monitor those investments. Even if the funding were available, some hedge fund managers might refuse to join an index platform. As a result, index providers have to strike a delicate balance between representativeness and economic efficiency.

32.10.2 Noninvestable Hedge Fund Indices and Five Complicating Issues

In theory, passive hedge fund investing could offer advantages over the FoF approach. Since indexing does not require research and monitoring costs, this eliminates the additional layer of fees and expenses charged by actively managed FoFs. Additionally, an investment in a hedge fund index should protect investors from manager selection risk. Just like indices for stocks or bonds, hedge fund indices deliver the normal returns of the asset class or investment style, which are referred to as hedge fund beta. By contrast, selecting individual hedge funds exposes investors to the risk of significantly underperforming the aggregate return of the hedge fund industry. Indeed, the dispersions of returns of both hedge funds and FoFs have historically been extremely high, and both exhibit a clear widening trend. A wrong fund selection decision could, therefore, have a large impact on performance.

Unfortunately, identifying a representative hedge fund index is not a trivial matter. The variety of index construction approaches and databases results in extreme heterogeneity of performances among these indices. For instance, Amenc and Martellini (2003) analyzed 13 different-style indices drawn from major index providers and observed a difference in performance of up to 22.04% in a single month for equity hedge indices. In addition, some indices that were supposed to measure the same strategy were negatively correlated to each other during certain periods. This is clearly confusing for investors and casts serious doubt on the possibility of using such broad-based hedge fund indices as yardsticks in performance measurement. But the challenges for passive hedge fund investors do not end there.

Gaining exposure presents another significant hurdle, as tracking broad-based noninvestable hedge fund indices is complicated by the following five issues:

1. Transparency: Many noninvestable hedge fund indices do not disclose the list of their components, their weights, or even their construction methodology. This significantly complicates the work of a third-party indexer unless the indexer uses privileged information from the index provider.
2. Unavailable funds: Most noninvestable hedge fund indices are partially made up of funds that are already closed to new investment or will be closed at some point in the future, once they reach their maximum capacity. A full replication by buying all the components in the index is often not feasible.
3. Liquidity: Traditional indexing approaches require regularly rebalancing a portfolio of hedge funds to minimize the tracking error with respect to some index. In many cases, this is not practicable because of the lack of liquidity of the underlying funds due to such constraints as lockups, redemption notice periods, and uneven cash flows in and out of funds.
4. Tracking error: Attempts to replicate the returns of noninvestable hedge fund indices by dynamically trading traditional assets such as stocks and bonds, or even futures and options, usually result in significant tracking errors, essentially because the target is an index of actively managed portfolios. Thus, although the

funds in the index do not change, the funds' individual securities and their key characteristics change continuously.

5. Reporting lag: Most broad-based hedge fund indices often report their net asset value with a considerable delay, such as three weeks after the end of the month. This means that a third-party index will always rebalance with a considerable lag.

32.10.3 Investable Hedge Fund Indices

The first investable hedge fund indices were launched in January 2001 by Zurich Financial Services and Schneeweis Partners and were soon followed by several others. All investable hedge fund indices share the goal of offering investors the opportunity to hold the entire hedge fund market at a relatively low cost. But as their construction rules, weighting, and rebalancing policies differ widely, significant variations in their performance result.

As mentioned previously, index providers face a trade-off between including more funds to be more representative and using fewer funds to facilitate index tracking. Indeed, to be investable, hedge fund indices must select only a limited number of liquid and open hedge funds. The process of selecting hedge funds based on certain criteria leads to adverse selection. That is, it is likely that the hedge fund managers who are willing to belong to an investable index would have characteristics different from those of the universe of all managers. Adverse selection is likely to lead to lower returns for the index, because it is often assumed that top-performing managers are less likely to agree to belong to an index. Most index providers impose strict selection criteria (e.g., minimum track record, minimum AUM, sufficient liquidity, absence of a lockup period, daily or weekly valuation, minimum transparency, willingness to accept additional investors, and commitment to provide sufficient capacity) in order to select the funds that are eligible to enter their index.

Several index providers go one step further and have signed partnerships with managed account platforms or have developed their own platforms in order to secure maximum capacity and liquidity on the components of their indices. Other providers have attempted to circumvent the problem altogether by using quantitative methodologies, which typically partition the hedge fund universe into clusters and use various algorithms to select the funds that are most representative of each cluster.

Regardless of the approach, the objective is to construct a representative sample of the hedge fund universe. As illustrated by Lhabitant (2007), the set of all investable indices included only four of the largest 25 hedge funds worldwide, despite the fact that those funds managed a total of more than $300 billion. Hedge funds with superior performance may find no need to be included in an investable index, as they are already capable of attracting investors, while less successful hedge funds may be more willing to comply with the strict selection criteria mentioned earlier, to increase their assets.

Beyond the question of representation, several criticisms of the selection process used by investable index providers have been raised. Since investable hedge fund indices are created with the implicit goal of launching a tracking vehicle, it is essential that their historical pro forma performance looks attractive to potential investors. Therefore, index providers have a tendency to select index members among funds with good track records. This does not, however, guarantee good performance in

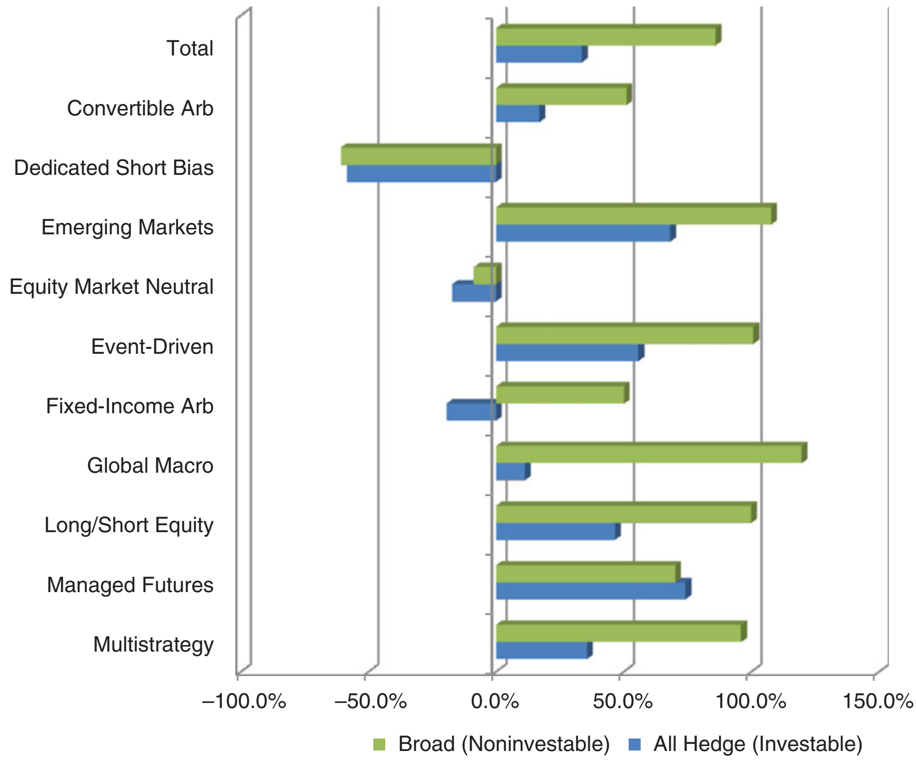


EXHIBIT 32.15 Total Return Achieved by Dow Jones Credit Suisse Investable and Noninvestable Indices from October 2004 to December 2014
Source: Dow Jones Credit Suisse.

the future. For most strategies and providers, a simple comparison between noninvestable and investable indices illustrates the underperformance of the investable index versus its noninvestable cousin (see Exhibit 32.15).

32.11 CONCLUSION

In the 1990s, funds of hedge funds were attractive to high-net-worth investors whose primary focus was on high returns. Beginning in the mid-2000s, institutions began to embrace FoFs as a way to invest in uncorrelated strategies, which served to lower their overall portfolio volatility. FoFs have always provided a valuable service by sourcing managers, performing due diligence, selecting strategies and funds, building portfolios, and providing risk management and ongoing monitoring of underlying funds. For this expertise, FoFs charge an extra layer of fees on top of what is paid to the underlying managers.

FoF performance has declined dramatically since 2000, experiencing a two-thirds drop in its 1990–99 annual return of roughly 11% to approximately 4%. Hedge fund performance has been the primary cause of this decline, with annualized returns dropping from roughly 18% in the 1990–99 time frame to roughly 6% since 2000.

A portion of this decline in returns may be attributed to lower interest rates and stock market returns, but the alpha earned by FoFs net of these market factors has, indeed, declined substantially. Fees continue to compress in the institutional markets as competent, lower-priced consultants enter the fray. FoFs continue to provide a valuable service for smaller investors who do not have the critical mass to properly diversify or the expertise to invest on their own.

Multistrategy funds provide a different type of access to investment talent in the form of a fund that is structured so that traders focus exclusively on their given talent without being distracted by operational or business issues. The MSF manager's role is primarily in allocating capital to the various teams, sourcing new trading teams, managing risk and leverage, providing a robust technology platform that can handle disparate strategies, and raising capital. While MSFs may promise investors a modest return in most markets since the MSF manager can tactically allocate capital quickly, it cannot be overemphasized that these funds, like most hedge funds that use leverage, can be exposed to periodic tail risk.

REFERENCES

Ackermann, C., R. McEnally, and D. Ravenscraft. 1999. "The Performance of Hedge Funds: Risk, Return, and Incentives." *Journal of Finance* 54 (3): 833–74.

Aiken, A., C. Clifford, and J. Ellis. 2014. "The Value of Funds of Hedge Funds: Evidence from Their Holdings." *Management Science* 61 (10): 2415–29.

Amenc, N., and L. Martellini. 2003. "The Brave New World of Hedge Fund Indices." Working paper, EDHEC Risk and Asset Management Research Centre, Lille, France.

Ang, A., M. Rhodes-Kropf, and R. Zhao. 2008. "Do Funds-of-Funds Deserve Their Fees?" NBER Working Paper W13944, April.

Black, K. 2006. "Improving Hedge Fund Risk Exposures by Hedging Equity Market Volatility, or How the VIX Ate My Kurtosis." *Journal of Trading* 1 (2): 6–15.

———. 2012. "An Empirical Exploration of the CBOE Volatility Index (VIX) Futures Market as a Hedge for Equity Market and Hedge Fund Investors." *Research in Finance* 28:1–18.

Brown, S. J., T. L. Fraser, and B. Liang. 2008. "Hedge Fund Due Diligence: A Source of Alpha in a Hedge Fund Portfolio Strategy." *Journal of Investment Management* 6 (4): 23–33.

Brown, S., W. Goetzmann, and B. Liang. 2004. "Fees on Fees in Funds of Funds." *Journal of Investment Management* 2 (4): 39–56.

Capocci, D., and G. Hubner. 2004. "Analysis of Hedge Fund Performance." *Journal of Empirical Finance* 11 (1): 55–89.

Darolles, S., and M. Vassie. 2010. "Do Funds of Hedge Funds Really Add Value: A Post Crisis Analysis." EDHEC-Risk Institute. http://faculty-research.edhec.com.

Fung, W., and D. Hsieh. 2002. "Hedge-Fund Benchmarks: Information Content and Biases." *Financial Analysts Journal* 58 (1): 22–34.

———. 2004. "Hedge Fund Benchmarks: A Risk-Based Approach." *Financial Analysts Journal* 60 (5): 65–80.

Kat, H., and S. Lu. 2002. "An Excursion into the Statistical Properties of Individual Hedge Fund Returns." Working paper, ISMA Centre, University of Reading, UK.

Lhabitant, F. 2007. "Delegated Portfolio Management: Are Hedge Fund Fees Too High?" *Journal of Derivatives & Hedge Funds* 13:220–32.

Lhabitant, F. S., and M. Learned. 2003. "Hedge Fund Diversification: How Much Is Enough?" *Journal of Alternative Investments* 5 (3): 23–49.

———. 2004. "Hedge Fund Diversification: Not a Free Lunch." In *Hedge Funds: Strategies, Risk Assessment, and Returns*, edited by G. N. Gregoriou, V. N. Karavas, and F. Rouah, 71–98. Washington, DC: Beard Books.

Liang, B. 2003. "Hedge Funds, Funds-of-Funds, and Commodity Trading Advisors." Working paper, Case Western Reserve University, Cleveland, OH.

Liew, J. 2003. "Hedge Fund Index Investing Examined." *Journal of Portfolio Management* 29 (2): 112–23.

Liew, J., F. Mainolfi, and D. Rubino. 2002. "Bifurcated Fund Analysis Model." *MFA Reporter*, November, 12–15.

Xiong, J., T. M. Idzorek, P. Chen, and R. G. Ibbotson. 2009. "Impact of Size and Flows on Performance for Funds of Hedge Funds." *Journal of Portfolio Management* 35 (2): 118–30.

CHAPTER 33

Hedge Fund Operational Due Diligence

This chapter provides an overview of the operational due diligence (ODD) process for hedge fund investors, building on the basic ODD process described in Chapter 10. It begins by distinguishing ODD approaches for hedge funds and private equity. The chapter then proceeds to discuss a variety of hedge fund–related operational risk considerations, including an analysis of trading procedures, cash management, service providers, compliance, and governance. The chapter also provides an overview of the operational decision and documentation process, and concludes by discussing considerations for performing an ODD review of funds of hedge funds.

33.1 DISTINGUISHING HEDGE FUND AND PRIVATE EQUITY OPERATIONAL DUE DILIGENCE

As has been outlined in other chapters, hedge fund investment styles are typically categorized into a number of different strategies, such as long/short equity, convertible arbitrage, and relative value strategies. While sharing some general commonalities, there are a number of differences related to the investment strategies employed. In general, however, we can classify most hedge funds as having two major distinctions from private equity funds that affect ODD.

First, most hedge funds trade more frequently than do private equity funds. This increased trading volume brings with it a number of accompanying operational practices and procedures that hedge funds are required to employ in order to support this more frequent trading activity.

Second, hedge funds typically trade, and hold, more liquid positions than do private equity funds. This deeper liquidity has consequences with regard to the way a fund approaches other operational areas, such as valuation. Enhanced activity in more liquid instruments, such as publicly traded companies, also has consequences on fund policies and procedures in other areas, such as compliance. While certain hedge funds may hold more illiquid positions, and some have been classified as private equity hybrids, the enhanced liquidity and trading volume of most hedge funds mean that investors in these funds need to tailor their ODD processes to more appropriately address the different risks involved.

33.2 FOUR OPERATIONAL STEPS IN ANALYZING HEDGE FUND OPERATIONAL TRADING PROCEDURES

In order to assess operational risk at a hedge fund, an investor must first collect and analyze operational data. As noted in the previous section, a key area of hedge fund operational risk relates to trading activities. The four operational steps in the hedge fund trading process are:

1. Execution
2. Posting and settlement
3. Allocation
4. Reconciliation

Execution refers to the process by which a fund completes a securities trade. To clarify, these securities do not necessarily need to be equities, as hedge funds trade a variety of positions along the liquidity spectrum. In practice, after the decision has been made to trade, an execution process would commonly begin. The trades would typically be communicated by investment personnel to a centralized trading desk. Some smaller hedge funds may not maintain dedicated trading desks; in these cases, trades are typically executed by investment personnel. In the context of an ODD review, the separation of investment and trading activities is considered best practice in order to lessen the potential for conflicts of interest.

Continuing our execution example, after routing trades to the trading desk, another consideration would be the method in which these trades are transmitted. In practice, physical trading tickets with time stamps are rarely, if ever, present these days. Instead, trading instructions are typically transmitted electronically through electronic messaging or an order management system (OMS). These software systems assist hedge funds in managing the trade life cycle process. The running list of all trades desired and completed during each trading day is commonly referred to as a **trade blotter**. After trading instructions are received by the centralized trading function, trades can then be executed with brokers or counterparties via either electronic interface or telephone. Hedge funds may record these telephone conversations in case there is a discrepancy regarding the actual executed trade and the instructions provided by the hedge fund. For electronic trading, different technology protocols, such as the Financial Information eXchange (FIX) protocol, may be used.

Posting is a term commonly used to refer to the process by which trades are logged internally at a fund, whether through order management or through fund accounting systems. **Settlement (internal)** refers to the process of reconciling third-party trade confirmations for executed trades with the firm's internal systems and trade blotters, and transferring the cash and securities to complete the trade. Depending on the types of securities processed, settlement procedures may occur practically in real time through automated reconciliations or may take several days to settle. In practice, the actual process of trade execution posting and settlement has multiple steps.

Trade allocation refers to the process by which trades are divided among the firm's various funds. It is considered best practice for a hedge fund to maintain a trade allocation policy that does not advantage one of the firm's funds at the expense of another. **Pro rata allocation** is a common allocation method by which a firm allots shares in the securities purchased to different funds based on predetermined proportionate amounts, such as assets under management, or a fund's predetermined desired allocation size.

After trades have been executed, posted, settled, and allocated among the firm's funds, the final step is reconciliation. **Reconciliation** refers to the process by which a hedge fund conducts another internal review to ensure that the internal details of the trade (buy/sell, security description, trade size, and price) are accurately matched with the details provided by the fund's counterparties. This reconciliation is typically performed as a **two-way reconciliation**, which is a reconciliation between the hedge fund trading records and the prime broker. Some hedge funds will use a third-party administrator. In those cases, a **three-way reconciliation**, or triangular reconciliation, is performed between the trading counterparties, the hedge fund itself, and the administrator. Due to the enhanced oversight provided by the addition of the fund administrator, it is considered best practice for a triangular reconciliation to be performed as opposed to a two-way reconciliation.

The timing of the completion of the reconciliation process depends on a number of factors, including whether a trade anticipated by the hedge fund failed to execute, which is commonly known as a **trade break**. The type of securities being traded may also influence the timing of the completion of the reconciliation process. For liquid securities, most reconciliations are typically completed on what is known as a **T+1 basis**, which means that one day after the trade date—referred to as *T*—the trade would be reconciled. Other securities, such as bank debt, do not generally settle as quickly and may not be reconciled for several days. An additional consideration relates to the timing of reconciliations. Although it is considered best practice to conduct frequent (i.e., daily) reconciliations, some hedge funds may opt to conduct less frequent reconciliations. Reasons for this include the increased operational work related to performing these reconciliations, as well as the potential additional expenses charged by fund administrators to conduct this work. However, longer times to settlement can actually increase the cost of trade errors, as errors will tend to be detected more slowly and the financial markets will have more time to move further away from the originally intended price.

During the ODD process, investors take measures to analyze the operational risks in each of the steps in the trade life cycle. One area this analysis focuses on is the independent oversight in place among different internal departments. For example, some hedge funds may employ fund accounting and operations departments with the designated task of supporting the post-trade life cycle process. In other cases, this process may be shared by individuals with multiple responsibilities, including those of managing the investment portfolio or for risk management. Due to the potential for manipulation and conflicts, it is considered best practice not only to segregate the investment and operational functions but also to balance separations among the various systems used in different functions of the trading process while still promoting process efficiency. That is, the tasks of portfolio management, trading, risk management, clearing, and valuation should be performed by different people.

Another area of consideration is the level of independent oversight of third parties in the trading process, such as the fund administrator or third-party valuation agents. For example, if a reconciliation is performed between the hedge fund and the administrator, and the lists of trades and settlement prices do not completely match, what happens next? Continuing our example, a fund administrator could independently work with the prime broker to analyze the discrepancy. Performing this research independent of the fund and the fund's traders would provide additional independence in this and similar situations. It is considered best practice for fund administrators to maintain internal procedures by which such situations are researched independently of the hedge fund. Although the fund administrator and the fund manager may maintain reasonable degrees of discretion, the reconciliation process should still be analyzed to highlight any material deviations from discretionary limits.

33.3 ANALYZING HEDGE FUND CASH MANAGEMENT AND MOVEMENT

The operational steps regarding trades just described relate to the movement of securities held by hedge funds. Related to these steps are a hedge fund's management and movement of cash. The way hedge funds account for and move cash is a key ODD consideration.

33.3.1 Four Primary Purposes of Hedge Fund Cash

The ways in which hedge funds deal with cash can be grouped into four primary categories:

1. Cash for fund expenses
2. Cash to facilitate trading
3. Cash flow to and from investors
4. Unencumbered cash

The operation of a hedge fund generates a myriad of expenses. These can include frequent recurring expenses, such as office rent and salaries, as well as less frequent expenses, such as audit and legal bills. In order to manage cash for fund expenses, there are a number of best practices hedge funds typically employ. One of these practices relates to the actual review of a vendor invoice. When an invoice for a fund expense is submitted to a hedge fund by a third party, it is considered best practice to have an individual who is not associated with approving payment of the invoice log the invoice into the firm's accounting systems and review it.

The reason for this should be fairly obvious. If an individual wanted to steal money from a hedge fund and he was in charge of both logging and paying invoices, all he would have to do is create a fake invoice and process the payment to himself. Segregation of duties seeks to mitigate this risk. Another method that hedge funds use to augment controls in this area is to have a preapproved list of vendors that are eligible to receive payments. If a new vendor pops up, such as the fake one in our example, then the new vendor would need to be added to the approved vendor list. Ideally, this vendor list would be subject to oversight by multiple individuals within the firm; thus, it would require a great deal of collusion to both add a fraudulent vendor to the approved vendor list and then log and process the fraudulent payment.

Another common practice with regard to the review of fund expenses at the hedge fund itself is to have multiple expense payment approvers from different departments throughout the firm. To provide an additional level of segregation, some hedge funds put together lists of senior individuals (i.e., an A list) and more junior individuals (i.e., a B list). Cash expense processing policies are then typically employed that require signatures from both an A list member and a B list member. Such differentiation adds another level of segregation to the process. Additionally, some hedge funds will go even further and segregate the number and type of approvers by the amount of the expense. Such a policy could be that all transfers up to and including $10,000 would require one A list signature and one B list signature for approval, but transfers of more than $10,000 would require signatures from two A list personnel and one B list member. It is also worth clarifying that during an ODD review, investors typically consider the actual form under which expense approvals are granted. These may be via a physical signature, via email, or through other electronic systems. Analyzing the approval process to ensure appropriate oversight and controls of approval channels is a key goal of ODD analysis.

The roles of the administrator and banks in overseeing the transfer of cash to process expenses are also critical for conducting an operational assessment of a fund's cash controls. In regard to fund expenses, after an internal review of the expenditure is conducted and approved by the fund, the next step typically involves the fund administrator, who will conduct a personal review of the expenditure. This review will typically focus on questions related to the amount of the expenditure and whether it is a common vendor that is frequently used. Checks performed by the fund administrator in this regard will typically also revalidate the hedge fund's internal approval process to ensure that the approvals were granted according to the hedge fund's policies. In sending over the fund's in-house approval work, it is also considered best practice for the fund administrator to be sent copies of the actual invoices received, so that the administrator may directly contact the vendor to determine the validity of the invoice or conduct additional inquiry regarding the expense if any further questions arise. After approvals are granted by the fund administrator, the final stage in the process is typically to transmit wire instructions to the bank in order to move the cash to process payment to the vendor.

The second primary category of cash is that used by funds to facilitate trading. We have addressed the process in which trades are executed and cash is transferred in order to make the purchase of a security by a fund. Depending on the type of transaction, hedge funds may also have positive cash balances on account with trading counterparties. A common example of a situation such as this arising would be with a swap counterparty. In these cases, a key consideration during the ODD process is to analyze what happens to this cash. One option would be for positive balances of cash to sit at the counterparty, waiting for the next trade to be executed. In other cases, hedge funds recall this cash, which is sometimes referred to as sweeping cash, an option that has two primary advantages for hedge funds:

1. *Reduction of counterparty exposure.* By taking this cash back from the counterparty, the hedge fund is minimizing its counterparty exposure. For example, if a trading counterparty becomes insolvent while it is holding the hedge fund's cash, the hedge fund would be unable to quickly recall this cash for trading.

2. *Option to reinvest or earn interest.* By sweeping back the cash from the counterparty to its own accounts, the hedge fund now has the ability to use the cash for trading or other purposes, such as to earn interest.

The frequency by which cash is swept varies among hedge funds. Some prefer the ease of leaving cash with a commonly used counterparty rather than sweeping back the cash and later depositing it again with the counterparty for future trading. Others may sweep cash daily, weekly, or monthly. It is generally considered best practice to sweep cash more frequently or when cash reaches a certain threshold amount.

The third primary category of hedge fund cash relates to cash used to process capital inflows and outflows, which are also called subscriptions and redemptions, respectively. Typically, the fund administrator is involved in processing the subscriptions and redemptions of investor funds subject to fund oversight. Similar to the movement of cash for fund expenses, a number of internal hedge fund review processes and administrator checks are typically in place regarding this capital. Such oversight is crucial for several reasons, including these four:

1. Ensuring that capital subscriptions are placed into the appropriate funds
2. Monitoring the process so that redemptions are not paid out to the wrong individuals and in incorrect amounts
3. Evaluating subscription and redemption timing to ensure that it complies with fund terms and timing guidelines
4. Reviewing subscription and redemption documentation to ensure that all appropriate procedures are in place for complying with regulations, such as anti-money-laundering and accredited investor provisions

The fourth primary category of hedge fund cash relates to cash that is not currently being used for trading but may be used in the future for either trading or another purpose; this type of cash is referred to **unencumbered cash**. Hedge funds typically earn interest on unencumbered cash by depositing the cash in liquid vehicles, such as checking accounts or interest-bearing money market accounts.

Best practices for ODD involve examining the procedures for managing cash with regard to each of the four purposes.

33.4 ANALYZING HEDGE FUND EXTERNAL PARTIES

Operational due diligence also involves contact with and analysis of external service providers. Chapter 10 on private equity operational due diligence discussed legal counsel and information technology in the context of ODD. This section focuses on prime brokers, hedge fund administrators, and external sources of information.

33.4.1 Analyzing Hedge Fund Prime Brokers

Due to the largely different trading strategies of hedge funds and private equity funds, Chapter 10 did not address the role of a common hedge fund service provider called a prime broker. **Hedge fund prime brokers** are institutions that facilitate hedge fund trading by aggregating a portion, or all, of a hedge fund's cash and securities as well as providing services for leverage and short selling.

As part of the ODD process, investors are increasingly focusing on solvency of prime brokers. After the collapse of Lehman Brothers, a firm that ran a large prime brokerage service among other businesses, hedge fund investors became increasingly concerned about prime broker solvency. In the case of the Lehman collapse, many hedge funds were unable to reclaim cash balances held at the institution in a timely manner. In part to diversify the risk of prime broker insolvency, it is now not uncommon for hedge funds to use two or more prime brokers.

It is also a goal of prime broker analysis to gain not only a further understanding of the solvency of the firm but also general background information. Investors seek the specific ways in which prime brokers may interact with the fund, in areas such as trade processing, ongoing counterparty management, and oversight, as well as with other service providers, such as fund administrators.

33.4.2 Analyzing Hedge Fund Administrators

Fund administrators perform two common functions: fund accounting and shareholder services. Unlike private equity funds, it is very common for hedge funds to use third-party administrators. Due to the additional oversight benefits of incorporating another party into the fund valuation process, virtually all investors today would no longer consider an investment in hedge funds whose funds are self-administered.

As part of their work, administrators are responsible for reviewing the valuations of securities held by funds. Typically, administrators are not responsible for independently valuing securities for which valuations are not readily available. These would traditionally be less liquid positions similar to those found in private equity funds. It is considered best practice for the hedge fund to provide the administrator with copies of the minutes of any internal valuation committee meetings as well as the pricing supports used by the hedge fund in valuing these positions. As noted in Chapter 10, in cases in which the manager directly prices a large portion of a fund's holdings, it is considered best practice for the fund to engage a third party, known as a valuation agent, to conduct an independent valuation of these positions.

For other securities that have readily available valuations, one key concern for investors is whether the administrator is able to independently verify the valuations provided by the hedge fund manager. For highly liquid positions such as public equities, this is not generally difficult, as a number of third-party pricing feeds are readily available. These valuation independence concerns, however, are particularly relevant for less liquid positions, such as certain over-the-counter (OTC) positions that are priced not through pricing feeds but through what are known as broker quotes. For broker-quoted positions, prices are obtained directly from brokers. So as not to rely solely on the valuation provided by one broker, it is typically considered best practice to obtain multiple broker quotes. When seeking to obtain valuations for broker-quoted securities, hedge funds will typically attempt to reach out to several brokers that deal in the security to obtain pricing quotations.

One way that administrators typically attempt to independently verify the prices submitted to hedge funds by brokers is through the creation of shared email in-boxes for pricing feeds. The way this works in practice is that when a broker sends a price to a hedge fund, the administrator is copied on the message. In this way, the administrator can accomplish the goal of independently collecting the prices sent to brokers, and can then use this data to independently verify the eventual pricing data submitted by the hedge fund. Not all administrators, however, may universally follow rigorous processes in this regard; therefore, it is up to investors to evaluate the robustness of the level of administrator oversight.

Another key concern for investors with regard to the valuation work of administrators relates to the administrator obtaining security prices from sources other than those used by the fund manager. The purpose of this is to provide additional diversification of pricing sources in order to enhance the accuracy of prices. By using multiple sources, less reliance is placed on the prices provided by one vendor or broker. As can be expected, in many cases the prices from these multiple sources may not exactly match. In these situations, different rules may be followed, including discarding large outlier prices and taking averages. Exhibit 33.1 outlines a typical comparison summary that would be produced by an investor during the ODD process to analyze the pricing sources used.

EXHIBIT 33.1 Example of a Hedge Fund and Administrator Pricing Source Comparison Summary

Instrument Type	Fund Pricing Policy	Internal Pricing Sources	Administrator Pricing Sources Used
Listed securities	Valued at last exchange sale price on the date or determination, or, if no sale occurred, valued at bid price (if held long) or ask price (if held short)	Third-party vendor feed A and third-party vendor feed B	Third-party vendor feed C
Listed options	Valued at the mean between the last bid and ask price	Third-party vendor feed A and third-party vendor feed B	Third-party vendor feed C
Over-the-counter (OTC) securities	Valued at last sale price, or, if no sale occurred, valued at bid price (if held long) or ask price (if held short)	Third-party vendor feed A and third-party vendor feed B	Third-party vendor feed D
Bank debt	Valued at last sale price or via broker quotes	Third-party vendor feed E or a minimum of two broker quotes	Minimum of three broker quotes, only one of which must be independent of the hedge fund

During the ODD process, a key consideration with regard to the work of fund administrators is also the timing of the production of the fund's net asset value (NAV). NAV is the value of the fund after liabilities have been subtracted. It is used in the production of investor statements to calculate the value of investors' holdings in the fund at different points in time. In practice, hedge fund NAVs are typically produced on a monthly basis, which is more frequent than the quarterly time line that is usually present for private equity funds. These statements are typically distributed within a number of days of the end of each month. For example, a hedge fund's statement of April's NAV might be sent to investors on May 7. The reasons for delays in the production of the final NAV often relate to delays in obtaining independent valuations for positions, which are sometimes known as marks. There are other reasons that may contribute to delays in NAV production, such as personnel turnover at either the hedge fund or the administration firm.

After the administrator and the hedge fund are both satisfied with the valuations obtained for a fund during a particular period (e.g., monthly), a final NAV is agreed on, which is known as cutting the NAV. Once the final NAV is produced, the administrator will prepare and distribute individual investor account statements. Exhibit 33.2 provides an example time line of the back-and-forth that goes on between the hedge fund and the administrator in the production of a month-end NAV.

From an operational risk perspective, continued delays in the production of NAVs can be representative of larger underlying problems between the hedge fund

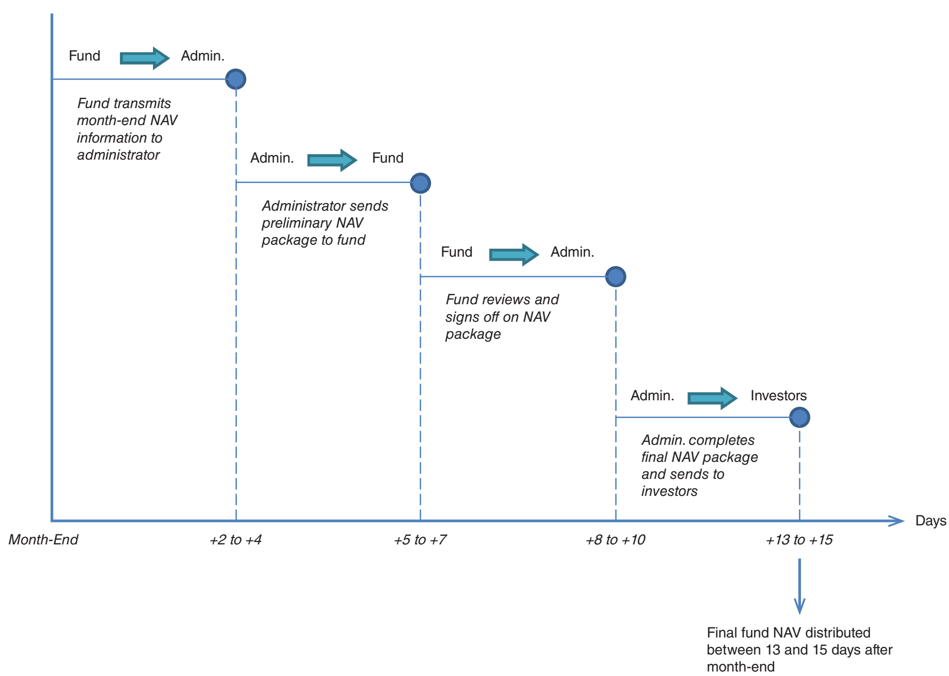


EXHIBIT 33.2 Example of a Hedge Fund and Administrator Month-End NAV Production Time Line

and the administrator. In this **operational signaling effect**, individual operational issues (e.g., continued NAV production delays) can signal that there are undiagnosed operational problems present. This effect can guide investors to focus more heavily on certain parts of their ODD. Continuing our example, if a hedge fund's NAV production has continually become more delayed each month, an investor should recognize that there is an operational signaling effect at play and should perform further due diligence on the specifics of the relationship between the hedge fund and the administrator in an attempt to further vet any undiagnosed problems.

33.4.3 Independent Service Provider Verification of Fund Operational Data

A core task of any ODD process is for investors to undertake an independent confirmation of key operational information. The goal of this analysis is for investors to not take a hedge fund's word at face value regarding certain key pieces of operational information, but to independently verify the fund's claims. In part as a result of the fraud undertaken by individuals such as Bernard Madoff, the two primary areas of focus relate to asset and position verification. **Asset verification** refers to the process by which an investor independently confirms a fund's level of asset holdings with third parties, such as fund administrators, prime brokers, and banks. The goal of such confirmations is to provide an independent comparison with regard to the asset level the fund manager is reporting and that which third parties are reporting. In practice, these asset levels may not tie out exactly. This is fine, as such reviews focus on the reasonableness of the information and not an exact audit of it. The goal would be to check, for example, that the fund manager is not reporting \$500 million under management when third parties report a balance of only \$200 million. These verifications should be performed before an investment is made, as well as on a regular basis for as long as the investor retains an allocation to the fund. **Position verification** refers to the process of confirming the holdings of actual fund positions with third parties, such as prime brokers and custodians.

33.5 ANALYZING HEDGE FUND COMPLIANCE CONSIDERATIONS

As part of an ODD assessment of a fund manager, there are compliance considerations that are also evaluated. The management of compliance risks within a hedge fund is typically overseen by a designated compliance department. Some smaller funds may not maintain dedicated compliance departments but instead have individuals who maintain multiple responsibilities, some of which are compliance related. There are three areas commonly overseen by the compliance department:

1. Initial and ongoing personnel training on compliance-related matters
2. Testing of the implementation of compliance policies
3. Monitoring and managing conflicts of interest

33.5.1 Personal Trading of Fund Employees

An example of a common hedge fund compliance policy on which training and testing are performed relates to the trading of securities by employees of the firm for their own accounts. This is known as personal account dealing, and the policies implemented in this regard typically apply not only to employees but also to their significant others as well as other immediate family members. One of the key concerns related to poor operational oversight of personal account dealing at a hedge fund is a process known as **front running**, or trading ahead, in which employees or others attempt to trade in advance of the firm's funds for their own accounts. Front running places the interests of the firm's employees ahead of its investors.

The majority of personal account dealing procedures define a universe of securities, known as covered securities, to which the policies apply. In practice, most hedge funds typically exclude from the universe of covered securities investments such as mutual funds, exchange-traded funds, and passively managed security accounts. The reason for these exclusions is the reduced likelihood that an employee would front-run anticipated hedge fund trades through these activities.

33.5.2 Common Compliance Risks Regarding Personal Trading

Pre-clearance of personal account trades is a process by which employees must seek approval from compliance before executing a trade. Pre-clearance requests may be submitted in a number of formats, including via email or through a designated personal account dealing management system.

Post-clearance refers to the process by which the compliance department collects employee brokerage statements and then attempts to reconcile them to pre-clearance requests. It is considered best practice for the compliance department of a hedge fund to collect employee brokerage statements directly from brokers and independent of the employee.

A **restricted list** is a list of securities that the firm has prohibited employees from trading. The reason a security is commonly placed on a restricted list is that the firm has received material nonpublic information regarding a particular security, which is commonly referred to as being conflicted out of a security. During the ODD process, investors seek to gauge who is responsible for maintaining this list and what controls are in place to prevent employees from trading in restricted names.

Minimum holding periods are requirements that prohibit an employee from purchasing a security and then selling it within a predefined period of time. The goal of such procedures is typically to prevent employees from spending too much time actively trading in their personal accounts in an attempt to garner short-term profits. In certain cases, hedge funds may employ a **hardship exemption** procedure, wherein an employee is allowed to sell a security at a loss even if it is within the minimum holding period. A hardship exemption is commonly allowed to limit exorbitant losses in employees' personal accounts.

33.5.3 Compliance Risks Regarding Insider Trading

Another common compliance risk consideration related to personal account dealing is **insider trading**, a scheme in which employees of a hedge fund use what is known as material nonpublic information (MNPI), sometimes referred to as insider information, in their investment process. Unfortunately, in recent years the hedge fund industry has been plagued by increased insider trading scandals.

In many countries, front running and insider trading are illegal and subject to criminal prosecution. In countries where there are no regulations regarding these activities, CFA Charterholders and CAIA Charterholders are required to abide by the *Standards of Practice Handbook*.

For example, Raj Rajaratnam of the Galleon Group worked with several industry coconspirators and was sentenced to 11 years in prison in October 2011 for insider trading. This was the largest hedge fund insider trading case in U.S. history, with total profits from the alleged scheme estimated at over $60 million.

The technical legal definitions of what constitutes MNPI can vary across jurisdictions. Many of the high-profile hedge fund cases that have been reported involving insider trading relate to the use of expert networks, which are organized by third-party for-profit firms. These networks are made up of groups of professionals and academics from various disciplines and industries that provide advice and consultations to hedge funds conducting research. A primary issue that recently arose with regard to the use of such networks is that some of the experts were either currently employed at or recently retired from publicly traded companies. Although these types of networks typically impose prohibitions on their experts with regard to communication of MNPI, some of the recent cases found that experts had communicated MNPI to hedge funds, and the hedge funds had subsequently—and illegally—profited from the information.

The use of such networks by hedge funds does not intrinsically represent a high degree of operational risk. Instead, the ways in which such networks are used and the associated compliance oversight are the determining factors in evaluating the level of operational risk associated with the use of such networks. Hedge fund compliance departments can take a number of steps to institute better controls over such networks, including these four practices:

1. Reviewing expert network policies regarding MNPI and ensuring that the policies comply with both applicable laws and the hedge fund's policies
2. Requiring preapproval from investment personnel before allowing conversations with experts
3. Having compliance personnel listen in on expert calls
4. Monitoring fund trading activity after expert conversations for unusual patterns

It should be noted that these same practices can be employed when third-party research sources that are not part of formal expert networks are used, such as freelancers. In addition to preventing intentional use of MNPI, compliance training can ensure that employees recognize and report when they believe they are in the possession of MNPI. At this point, the firm can add the particular security to the restricted list.

33.5.4 Electronic Communication Monitoring

Hedge funds use electronic communications to facilitate their daily investment and operational procedures. Electronic communications include email, instant messaging, and remote device communications. Some jurisdictions require hedge funds to archive electronic communications so that they may be researched in the event that any questions relating to trading activity arise. One of the key reasons for this is to monitor for and prevent the transmission of MNPI, which could facilitate insider trading. During the ODD process, investors attempt to gauge the level of oversight by the firm's compliance department with regard to actively monitoring these communications across all devices. In certain cases, hedge funds maintain specialized systems that facilitate this ongoing monitoring. Monitoring can take place via a number of methods. Two common methods are:

1. Lexicon-based searches that locate messages containing keywords or phrases that may signal MNPI and warrant further investigation
2. Perusing a randomly selected fixed percentage of firm-wide and department-wide emails

33.5.5 Analyzing the Work of Third-Party Compliance Consultants

The area of compliance risk management has become increasingly complex. This has been fueled by a number of regulatory initiatives, such as the Dodd-Frank Act in the United States and the Alternative Investment Fund Managers Directive (AIFMD) in Europe. To ease the burden on internal hedge fund compliance departments, hedge funds may sometimes work with third-party compliance consultants. These consultants may perform a number of tasks, including these four actions:

1. Performing mock regulatory audits
2. Updating compliance policies and procedures to comply with new regulations
3. Assisting with, or completely running, compliance testing programs
4. Assisting with regional compliance expertise in satellite offices where the firm may have little to no internal compliance personnel

It is generally considered best practice for a hedge fund to use a third-party compliance consultant to augment the work of the internal compliance function. Investors will commonly review the work product of these consultants as well as analyze the ways in which hedge funds may have implemented the compliance changes proposed by these firms in order to gauge how effectively the firm is monitoring compliance risks.

33.6 DOCUMENTING THE OPERATIONAL DUE DILIGENCE PROCESS

Chapter 10 introduced the eight core elements constituting the private equity ODD process. Several stages of the ODD process could be classified as being different aspects of the larger data-gathering exercise. These stages lead to the analysis of the operational data collected. The analysis of data typically results in not only the logging of notes and events related to the review but also the production of a document summarizing the analysis. This is the sixth core element in the ODD process, known as process documentation. Here the aspect of process documentation that we focus on is the creation of a document detailing the results and conclusions of an ODD investigation.

Individual organizations approach the documentation process in different ways. Some organizations may produce a brief summary memo that outlines only the key concerns from the ODD process. While such summary memorandums are efficient for review, it is often considered best practice to produce a more detailed document. The length of ODD process documentation reports has grown to keep up with the broadened scope of ODD reviews.

These detailed documents typically begin with an executive summary section, which summarizes the key findings from the ODD review. It is also considered best practice not only to document concerns in the executive summary section but also to detail fund and firm strengths uncovered during the ODD process. The other sections of the report typically provide detailed analysis of each of the operational risk review areas covered during the ODD process. One of the key benefits of producing a detailed report summarizing ODD analysis is that in the event that an investment is made with a fund, the document facilitates the ongoing monitoring process. Should an investment not be made initially, the detailed ODD report can be used to facilitate a revised review of the same manager if the fund is still of interest to investors in the future.

33.7 OPERATIONAL DECISION MAKING AND ALLOCATION CONSIDERATIONS

The end result of the ODD process is to facilitate the development of an operational risk determination, sometimes referred to as an operational decision, which is the seventh element of the process introduced in Chapter 10. An operational decision by an investor, or investment organization, can typically result in a number of common allocation conclusions, as summarized in Exhibit 33.3.

EXHIBIT 33.3 Common Allocation Conclusions and Resulting Actions from the Operational Due Diligence Process

Conclusion	Common Possible Resulting Actions
High level of operational risk that is unacceptable to investors	Decision is made to make no investment with the hedge fund initially. The decision may be reevaluated to make an allocation in the future if the fund makes operational improvements.
Intermediate level of operational risk that is acceptable, but still concerning to investors	A reduced allocation is made to the hedge fund.
Low level of operational risk where operational risks may still be present but classified as relatively low risk by an investor	An allocation of the original amount anticipated is made.

In reference to Exhibit 33.3, it should be noted that in each case, an investor can share with the hedge fund under review feedback relating to the operational risk concerns discovered. This feedback can then be used by the hedge fund to repair operational deficiencies, with the goal of securing an initial or increased allocation from an investor in the future.

It should also be noted that the guidelines in Exhibit 33.3 are simply general guidelines. Individual investors may make different determinations relating to the severity of certain operational deficiencies. This relates to a key operational risk concept known as factor weighting, which refers to the importance (i.e., weight) that individual investors give to different operational risk considerations when coming to an overall operational decision. For example, some investors may choose to equally weight the various operational risk factors analyzed. Other investors may view certain items, such as fund accounting and valuations, as holding more operational risk, and therefore give those areas more weight when making operational decisions. While there is flexibility in assigning factor weights, there is a series of market norms regarding what is considered operational best practice in certain areas that guide investors in the factor weighting process. **Operational benchmarking** is the process of comparing operational best practices to the actual procedures in place at a fund. Operational best practices may vary by strategy, as a macro fund trading futures and swaps may have different procedures and regulations than a long/short equity fund has. It should also be noted that the factor weighting process is not necessarily quantitative in nature; qualitative weights, such as designating a particular operational risk area as "very important," may also take place.

There is a general consensus that certain practices are no longer acceptable to the majority of investors. An example of this would be a fund that self-administers its own funds. This doesn't mean that no investor would ever invest in a hedge fund that self-administers but simply that it is no longer the norm. Investors who oppose self-administration often feel that this issue is so important that a hedge fund that self-administers would automatically be precluded from investment eligibility, regardless of anything else that is going on at the fund. Said another way, they give the area of self-administration a high factor weighting, and if a hedge fund chooses to self-administer, it is automatically eliminated from being eligible for investment. This yes-or-no approach to analyzing operational risk in different areas is known as a binary operational decision process because there are only two outcomes.

In the example, the issue of self-administration would be referred to as an **operational threshold issue**, meaning that it is an issue that must be satisfied in order to have a particular investor continue to consider allocating to a particular fund. If the hedge fund is above an investor's operational threshold (i.e., does not self-administer its own funds), then it remains within the investable universe. To be clear, investors typically do not view each operational risk area to be a threshold issue. For example, most investors would apply a degree of flexibility to the number of approvers required to sign off on cash transfers. Other common examples of operational threshold issues for many investors would be a hedge fund that does not use a Big Four accounting firm or that employs many relatives in the same department.

In documenting ODD reports, some investors attempt to assign quantitative operational risk scores to the different operational risk areas covered. These scores can then be weighted based on predetermined factor weightings in order to produce a weighted average operational risk score. While such approaches may be useful for high-level operational risk analysis or comparing scores among different hedge fund managers, it is generally considered best practice to use more of an information-based approach to operational risk analysis and to focus on the actual operational practices employed rather than just scores.

33.8 INVESTIGATIVE DUE DILIGENCE

Investigative due diligence, sometimes known as background investigations, was traditionally viewed as being a distinct part of the due diligence process. In practice, however, background investigations were typically coordinated as part of the larger ODD process. Reasons for this coordination included that such examinations typically do not directly address the investment process.

33.8.1 Three Models of Selecting Personnel for Investigation

Investors traditionally perform background investigations on key fund personnel as well as on the management company entity for the fund. There are three common approaches investors use to assist them in determining which individuals to investigate:

1. EQUITY OWNERSHIP MODEL: Under the **equity ownership model** approach, an investigation would be performed on all personnel who have equity ownership in the management company of the hedge fund organization. This is generally feasible from a cost and investigation-duration perspective for a small hedge fund organization with two or three owners. In larger hedge fund organizations, which typically have a large number of owners, such searches can become prohibitively expensive and lengthy.
2. INVESTMENT DECISION-MAKING AUTHORITY MODEL: The **investment decision-making authority model** approach focuses on performing background investigations on those individuals who have authority to make investment decisions and act (i.e., trade) on such decisions. This generally includes portfolio managers and traders. Rather than focusing on key operational personnel within a firm, this model focuses on vetting the backgrounds of those involved with the investment process on a daily basis.
3. RISK CONTROL MODEL: Under the **risk control model** approach, background investigations are performed on all individuals, both investment and noninvestment focused, who control risk within an organization. These include portfolio managers and traders, as well as senior operational personnel, such as the chief financial officer and chief compliance officer. Depending on the organization under consideration, this model generally encompasses the most individuals, as compared to the other two models.

It should be noted that there are no universal rules with regard to which individuals should be investigated. During the course of an ODD review, an investor may learn that an individual with a seemingly unimportant title may turn out be a key individual in the organization's operations. In these cases, flexibility in the approach employed to select individuals for investigation is often merited.

33.8.2 Five Areas of Background Investigation

Five areas are commonly covered during the background investigation process:

1. Criminal searches
2. Civil searches
3. Regulatory searches
4. Media searches
5. Factual information searches

Criminal searches involve reviews of all records related to criminal activities. Areas typically covered during a criminal search include the following:

- Arrest records
- Parole and probation records
- Criminal conviction records
- Traffic and moving violations

Civil searches cover all civil litigation records, including the following:

- Court dockets
- Bankruptcy records
- Foreclosures
- Tax cases
- Other judgments and liens

Regulatory searches involve review of the following:

- Hedge fund regulatory filings
- Records of financial regulators, such as the Securities and Exchange Commission (U.S.) and the Financial Conduct Authority (UK)
- Office of foreign asset control and global sanction searches

Media searches include the following:

- Internet-based searches of web content
- Broad media searches of industry periodicals
- Social media screens

Factual information searches include the following:

- Confirmation and review of previous employment and educational endeavors
- Searches of asset records, such as property and vehicles
- Searches of fictitious-name databases
- Review of other business interests outside the hedge fund
- Reference checks

33.8.3 Factors in Organizing and Interpreting an Investigation

Depending on where these investigations are performed, different levels of detail may be available. This is in part due to differing privacy laws across countries, as well as to the different structures of the previously noted areas of investigation, such as court system structures. Additionally, in places where the primary language differs from the primary language of the investigators, difficulties may arise. In most cases, hedge fund personnel are required to sign releases in order to provide permission for institutions to release certain information about a person's previous affiliation with the institution.

Traditionally, most investors simply outsourced the background investigation process to third-party investigative firms. This approach suffered from a separation between the investigative and the ODD process. As a result, it is increasingly common for investors to combine the background investigation and ODD processes under a single entity, such as a third-party ODD consulting firm. Under these integrated reviews, investors benefit from more comprehensive deep-dive due diligence reviews and can subsequently make more informed investment decisions.

Evaluating the results of the investigative process can be a very subjective matter, on which reasonable minds may disagree. For example, if a hedge fund manager had been previously convicted of financial crimes, most investors would agree that this would be a deal breaker, and they would not invest with this manager. Still, depending on the nature of the offense, some investors might choose to be more forgiving. Outside of directly financial-related matters, consider the situation of a hedge fund manager who is charged with driving while intoxicated but pleads guilty to a lesser charge, such as reckless driving. Individual investors might evaluate such circumstances differently when coming to an overall investigative due diligence assessment of a hedge fund.

Additionally, it is becoming increasingly popular for investors to conduct ongoing investigative due diligence on hedge fund management companies and key personnel. These ongoing investigations typically focus on items that may change over time, such as arrest records and civil litigation. In these cases, investors follow different approaches with regard to the scope and timing of such searches; however, the goal is to ensure that some degree of ongoing investigative due diligence is performed, rather than having the initial review be the final source of information.

33.9 FOUR APPROACHES TO RESOURCE ALLOCATION FOR OPERATIONAL DUE DILIGENCE

In designing their approach toward allocating resources to conducting ODD reviews of funds of all types, including hedge funds, investors typically select one of four popular approaches. These approaches are dedicated, shared, modular, and hybrid. A definition of each of the style categories follows:

1. DEDICATED: A **dedicated operational due diligence approach** is an ODD framework in which an investment organization, such as a fund of hedge funds, has at

least one employee whose full-time responsibility is vetting the operational risks of hedge fund managers.

2. Shared: A **shared operational due diligence approach** is a framework in which the responsibility for reviewing the operational risk exposures at hedge funds is shared by the individuals who have responsibility for investment due diligence. No full-time, dedicated ODD staff are employed.
3. Modular: A **modular operational due diligence approach** is one whereby the ODD process is classified into functional components and divided among specialists with relevant domain-specific knowledge. It is important to note that in a modular ODD framework, domain experts typically have responsibilities in addition to their ODD duties. Some of the titles that these domain experts typically hold within a fund of hedge funds organization include general counsel and chief financial officer. Under a modular approach, these domain experts typically collaborate their ODD work through the leadership of an individual, whom we will refer to as an operational generalist.

   The operational generalist can be thought of as an information aggregator, who pieces together the functional reviews completed by all the domain experts to facilitate the investment organization moving toward an operational risk conclusion. The ODD duties of the operational generalist can be very similar to those of ODD analysts under a dedicated framework and can include such things as on-site manager visits and operational risk report generation. Under a modular framework, the operational generalist (or a group of individuals performing the operational generalist function) can either be a dedicated ODD professional (i.e., fitting into the definition of the dedicated approach) or serve other functions within the organization, and may even be a domain expert in her own right. An example of this would be a chief financial officer who has other responsibilities within the investment organization yet who also serves as the operational generalist piecing together the ODD work of the domain experts.
4. Hybrid: A **hybrid operational due diligence approach** refers to an approach that encompasses some combination of the three previously described approaches (i.e., dedicated, shared, and modular). An example of a hybrid framework would be an investment organization that employs a full-time ODD analyst (dedicated framework) while leveraging off in-house domain experts as needed (modular framework). Continuing this example, these domain experts would not be a part of the standard ODD review process followed by an investment organization (such as under a modular approach) but used on an ad hoc basis. Another example of an ODD framework that would fall under the hybrid classification would be an investment allocation organization, such as a fund of hedge funds, that outsources the ODD function, either in part or entirely, to a third-party ODD consultant. Therefore, within those managers that fall into the hybrid classification, it is important to note that a significant diversity of approaches may exist.

It should be noted that each of these ODD style buckets refers to the framework implemented by an investor or investment organization that performs ODD reviews. These style buckets do not address which individuals or groups at a fund of hedge funds holds the authority to make the ultimate operational conclusion regarding a particular hedge fund manager. Furthermore, these style buckets do not address which individuals or groups, such as an investment committee, have the authority to make the final allocation decision to a hedge fund manager, which combines the views of the operational and investment due diligence processes.

Due to the scope and complexity of operational risk reviews, it is increasingly popular for funds to employ some mix of either dedicated or hybrid approaches. Under hybrid approaches in particular, there has been an increasing use among investors of third-party ODD consultants, which seek to augment the in-house ODD work of investors.

33.10 HEDGE FUND GOVERNANCE

Fund governance can be defined as an interconnected system of controls and procedures that determine independence, transparency, and oversight throughout the fund. Analyzing fund governance structures is essential to conducting a thorough assessment of a hedge fund's operational risks. Governance can be implemented through the work of investors and service providers, and within the hedge fund itself.

One common governance structure in place at hedge funds is internal committees, which are responsible for providing oversight and transparency across a number of investment and operational areas. There are five common operational hedge fund committees:

1. Operations committee
2. Valuation committee
3. Business continuity and disaster recovery committee
4. Best execution committee
5. Compliance committee

Another common hedge fund governance mechanism is a fund's board of directors. The **board of directors** is a group of individuals who are responsible for fulfilling regulatory obligations, exercising legal rights, and providing limited independent oversight of funds. These individuals may be stand-alone professionals or may work for third-party professional directorship companies that specialize in providing directors for hire. From a U.S. perspective, it is common to focus on the board of an offshore fund located outside the United States.

Outside the United States, the term *offshore funds* could also refer to funds that are not located in the investor's home domicile. Not every offshore jurisdiction necessarily requires funds to maintain boards; however, popular offshore hedge fund jurisdictions, such as the Cayman Islands, require boards to be present. The exact duties and rights of directors are typically designated in the fund offering memorandum as well as in accompanying documents, such as the articles of association. Common duties of offshore board members include these six:

1. Overseeing the enforcement of any **redemption gates**, which is a mechanism by which a hedge fund can limit the total amount of capital being redeemed from a fund at any single redemption period (e.g., quarterly). During the financial crisis of 2008, many hedge funds decided to enact hedge fund redemption gate provisions, a process known as lowering the gate. Although practically any decision

to lower the gate would be made in consultation with the hedge fund itself, in many cases it would technically be the decision of directors to lower the gate for certain funds whose boards they served on.

2. Reviewing and approving the audited financial statements of a fund.
3. Approving any amendments to legal documentation.
4. Approving the hedge fund manager's use of certain mechanisms or altering the original terms of the mechanisms, such as an audit holdback and reserves. An **audit holdback** is a mechanism by which a hedge fund manager may retain a portion of an investor's redeemed capital until the finalization of a hedge fund's audit. The reason for these holdbacks is to provide a capital buffer to the fund manager should the financial figures be different from expectations. A **reserve** is a general term describing a capital buffer that can be created by a hedge fund from invested capital to meet potential expenses. The reasons for creating reserves can vary greatly. Examples of such situations would be creating a reserve in advance of litigation or other unusual expenses.
5. Reviewing fund manager valuations and overseeing the enforcement of valuation practices and procedures.
6. Reviewing the ongoing performance of fund service providers and approving new fund service provider appointments.

To carry out their oversight, directors typically engage with fund managers on at least a semiannual basis. While it is considered best practice for hedge funds and directors to conduct at least some of these meetings in person, the interaction may also take place by telephone or videoconference. In practice, some directors engage with fund managers more frequently to facilitate enhanced oversight. Due to the relatively infrequent nature of director and hedge fund interaction, directors are not charged with overseeing the day-to-day functions of the hedge fund. Instead, they place the responsibility for this day-to-day oversight with other parties, such as the hedge fund itself, and other service providers, such as fund administrators. It should be noted that this delegation of duties does not remove directors' responsibility to provide oversight but allows them to rely on the work of those who are more involved on a day-to-day basis with the hedge funds.

In general, the boards of these offshore funds must consist of a majority of individuals located in the offshore jurisdiction in which the fund is domiciled. For example, the common makeup of an offshore fund board could be two independent directors and one individual from the hedge fund. Due to the general requirements for local offshore residency for independent directors, a common consideration with regard to offshore directors relates to the notion of **director capacity**, which refers to the maximum number of board positions that a director may serve on in order to maintain effectiveness. In extreme circumstances, some directors have sat on upward of 100 boards. Directors who sit on large numbers of boards are referred to as jumbo directors. Many investors, questioning the ability of a director to function effectively when sitting on very many fund boards, have proposed more modest limits.

Another concern for investors is the use of **associate directors**, who support the work of the individual named to serve as the fund director. Associate directors are also sometimes called supporting or junior directors. The individuals named to serve as directors, often designated in the hedge fund offering memorandum, are sometimes called **formal directors** or official directors. Associate directors are typically more junior and less experienced than formal directors. The use of associate directors is more common at larger directorship firms. Although the role of associate directors can vary among organizations, they can typically perform all the tasks of formal directors, including reviewing information from the hedge fund, drafting director reports, and even attending meetings in their place. There is nothing inherently wrong with the use of associate directors, but due to their ability to effectively step into the shoes of formal directors, investors are increasingly focusing on the qualifications and duties of associate directors alongside those of formal directors.

In addition to reviewing the qualifications of directors, investors are increasingly requesting background investigation reviews as part of their investigative due diligence process. The scope of these investigative due diligence reviews is typically the same as would be employed for senior personnel of the hedge fund itself.

Additionally, investors are increasingly reviewing the operational practices in place at directorship companies. Reviews of these practices typically involve evaluating the quality and appropriateness of many of the same areas reviewed when appraising other hedge fund service providers:

- Business continuity and disaster recovery plans
- Compliance infrastructure
- Type and amount of insurance coverage
- Information technology infrastructure
- Scalability of existing operational infrastructure

In certain cases, hedge funds may employ advisory boards that effectively oversee several areas of the firm, including the work of the offshore board of directors. While the duties of these boards can vary, there are three main benefits of such boards:

1. Additional layer of oversight.
2. Potential to be engaged on a more frequent basis with the hedge fund than the offshore board.
3. Ability to staff a board with individuals having more diverse skill sets or experience that is directly relevant to the hedge fund strategy under review as compared to the offshore fund boards; an advisory board may include representatives from large external investors, who can give additional points of view on investment issues.

33.11 HEDGE FUND INSURANCE

A key area that is often overlooked as part of the ODD process is the level of insurance coverage in place at the hedge fund. Like all businesses, hedge funds are required to carry certain amounts of insurance. Some types of insurance held by hedge funds are general in nature and not necessarily specific to the hedge fund or financial industry. Other types are more appropriately customized to the work of the hedge fund.

During an ODD review, it is important to analyze the types and amounts of coverage in place to ensure that they are appropriate. A hedge fund that is underinsured may be overly exposed to risk of loss. Similarly, a hedge fund that is overinsured may be too cautious or managing its internal available capital resources inappropriately by spending too much on insurance. Additionally, a hedge fund that has overlooked certain types of insurance coverage may be ignoring key risk areas. Finally, the types of insurance coverage in place may not only protect the hedge fund from loss but extend to investors and other service providers as well.

The common types of insurance coverage maintained by hedge funds include errors and omissions (E&O), directors' and officers' liability coverage (D&O), general partner liability coverage, and employment practices liability coverage. In some cases, certain types of coverage may be segregated into smaller insurance policies or referred to by other names. An example of this would be a general partner liability policy being split into separate management liability and professional liability policies. Some hedge funds go beyond this basic level of coverage and obtain other types of policies, such as key-person insurance and business personal property and general aggregate liability policies. During the ODD process, investors typically inquire about the insurance in place at fund service providers, much of which is similar to that maintained by the hedge funds themselves, such as general business coverage.

33.12 PERFORMING OPERATIONAL DUE DILIGENCE ON FUNDS OF HEDGE FUNDS

In certain instances, an investor may seek to allocate capital to a fund of hedge funds, which is an organization that allocates capital to underlying hedge funds. It is considered best practice to conduct an ODD review of the fund of hedge funds as part of the overall due diligence process. When conducting an ODD review of a fund of hedge funds, it is important to consider the many differences between the operational procedures employed at a hedge fund and those performed at a fund of hedge funds.

The most common difference relates to the fact that the fund of hedge funds typically does not execute trades directly in the market but rather allocates capital to underlying hedge funds. In these instances, investors would focus on the ways these capital transfers are carried out. For example, which individuals are allowed to approve the actual transfer of capital to the underlying funds? In the context of an ODD review, the analysis of the capital subscription and redemption process at a fund of hedge funds is similar to the analysis of a hedge fund trading process.

When analyzing a fund of hedge funds, as opposed to a hedge fund, it is considered best practice to conduct an evaluation of the quality of the fund of hedge funds' ODD process. Funds of hedge funds should be performing ODD and investigations on the underlying hedge funds to which they allocate. Consequently, investors should take measures to evaluate the resource allocation, scope, quality, and robustness of the fund of hedge funds' ODD and investigative processes as well.

Investors in funds of hedge funds should also consider liquidity issues. When a fund of hedge funds makes an investment in an underlying hedge fund, it is subject to the lockup and redemption policies of that hedge fund. If the underlying hedge funds have an average lockup period of one year with quarterly redemptions, whereas the fund of hedge funds has a six-month lockup period with monthly redemptions, this liquidity mismatch can cause trouble for investors when large redemption requests are made.

33.13 CONCLUSION

Hedge fund investors need to conduct operational due diligence before allocating assets to a hedge fund and must continue to follow developments at each hedge fund over the life of their investment. Investors should be concerned about a wide variety of potential issues during the ODD process, including trading, valuation, and settlement procedures and custody, prime brokers, and cash movement processes, as well as the ethics, compliance, independence, and experience of the fund's employees.

Operational due diligence for hedge funds differs from ODD processes for private equity funds, as hedge funds trade more frequently and hedge fund investors have much more frequent opportunities to redeem their investments. ODD is often performed by a different team of analysts than those responsible for investment due diligence. Under a modular due diligence process, many employees of the investor's firm will become involved, such as those experts in compliance, systems, and accounting. The duties of hedge fund directors, an important role in offshore jurisdictions, were also discussed in the chapter.

CHAPTER 34

Regulation and Compliance

Whether we are reviewing the financial regulatory system of the United States, Europe, or Asia, there are certain principles and tensions that inform the regulatory scheme of free market economies. Investors and financial industry professionals occupy a world of regulation and compliance. For those in this environment, it is important to understand that regulations are a compromise between and among governments and business interests as well as social and monetary policies.

34.1 THREE FOUNDATIONAL PRINCIPLES OF FINANCIAL MARKET REGULATION

Regulatory oversight and enforcement are the way governments and political entities ensure adherence to these compromises and policies. Because they exist in a political environment of compromise, enforcement of these regulations is subject to change through both interpretation and implementation. By example, the recent prosecution of individuals in the United States for insider trading and the United Kingdom's pursuit of large financial institutions for criminal charges relating to trading and collaboration are based on regulations and laws that have been on the books for some 70 years, although seldom used due to resource considerations or systemic concerns.

Beyond the individual dynamics of a given political jurisdiction, capital markets are built on trust. Financial regulations are based on principles of (1) transparency, (2) market integrity or fundamental fairness, and (3) government protection of economic and social systems through the rule of law. Although regulations will change from country to country, these guiding principles will provide the investor or industry professional with the beginning point of analysis in most situations. It is important to note that financial regulations are laws, and laws often carry both pecuniary and criminal sanctions for those who violate their terms, so it is never enough to comply with the letter of regulations.[1] We must go further.

While reviewing this chapter or using its content to make investment or business decisions, the following must be firmly understood: (1) this material is dated (remember that the regulatory environment is constantly changing); (2) all regulations are subject to interpretation, and thus work must be done to ensure that one understands how and why regulators are interpreting regulatory provisions at the moment; (3) there are no global rules, so one must pay very close attention to and observe the rules of the applicable jurisdiction, never assuming to know the rules but always checking and verifying; and finally, (4) most financial regulatory schemes provide for civil, monetary, and/or criminal penalties against individuals and institutions for noncompliance or violations. These penalties are codified in governmental regulatory/administrative actions, treaties, and judicial opinions, and may be stated within the relevant statutory framework. For the most part, the penalties cover a wide spectrum of activities, and their severity is meant to correspond to the perceived harm done to the financial system or to the actual harm done to specified victims. History has shown that the implementation of penalties is fluid, can be inconsistent, and varies depending on such factors as the quality of legal representation, factual scenarios, and the disposition of decision makers. What the reader should take away from this discussion is that it is almost impossible to state with any meaningful degree of certainty the risks associated with violations of the securities laws and regulations. Thus, it is important to ensure that your business operations are compliant with local laws, whatever the jurisdiction.

Throughout this chapter, the focus will be on alternative investments. These investments are construed as being not subject to the typical auction system and daily pricing of exchange funds or securities. While virtually every structured product can be deemed an alternative investment, the line for instruction has to be drawn somewhere. It should also be understood that, as this chapter reviews the regulatory structures of numerous political and legal jurisdictions, it cannot, by definition, be exhaustive. Thus, the chapter is designed to provide basic building-block information, which the reader can build on when conducting further analysis. The chapter begins with the United States, moves on to Europe, and finishes with selected Asian countries.

34.2 THE REGULATION OF ALTERNATIVE INVESTMENTS WITHIN THE UNITED STATES

The starting point of financial regulation in the United States is the **Securities and Exchange Commission** (**SEC**), which is an independent agency of the United States federal government and has primary responsibility for the enforcement of all federal securities laws and the regulation of the securities industry.[2] The SEC derives its authority from the **Securities Exchange Act of 1934**, which provides governance of securities transactions on the secondary market (i.e., after the initial public offering) and regulates the exchanges and broker-dealers in order to protect the investing public. All companies listed on stock exchanges must follow the requirements set forth in the Securities Exchange Act of 1934. Primary requirements include registration of any securities listed on stock exchanges, disclosure, proxy solicitations, and margin and audit requirements. Besides the Securities Exchange Act, the SEC is tasked with enforcing the Securities Act of 1933, the Investment Company Act of 1940, the Investment Advisers Act of 1940, and the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010.

34.2.1 Legal Foundation of Regulation of Hedge Funds in the United States

The **Securities Act of 1933** has two main goals: (1) to ensure more transparency in financial statements so investors can make informed decisions about investments, and

(2) to establish laws against misrepresentation and fraudulent activities in the securities markets. Before this legislation, the sale of securities was primarily governed by state laws; however, the market crash of 1929 raised serious questions about the effectiveness of the way markets were being governed. The **Investment Advisers Act of 1940** defines the role and responsibilities of an investment adviser. Until 1996, most investment advisers were subject to regulation by both the SEC and one or more state regulatory agencies. The Act was amended in 1996 and again in 2010 to allocate regulatory responsibility between the SEC and the states. Today, most small advisers and mid-sized advisers are subject to state regulation of advisers and are prohibited from registering with the SEC. Most large advisers must register with the SEC, and state adviser laws are preempted for these advisers. The act also identifies investment advisers' liability and provides guidelines regarding the fees and commissions they can collect. Finally, the **Dodd-Frank Act** is a set of federal regulations passed in 2010 in an attempt to prevent the recurrence of events that caused the 2008 financial crisis. Primarily affecting financial institutions and their customers, the act is designed to lower the risk in various parts of the U.S. financial system. It established new government agencies, such as the **Financial Stability Oversight Council** and the **Orderly Liquidation Authority**, which monitors the performance of companies deemed "too big to fail" in order to reduce the potential for the kind of widespread risk seen during the 2008 financial crisis. The new **Orderly Liquidation Fund** provides money to assist with the liquidation of financial companies that have been placed in receivership because of their financial weakness. Additionally, the council can break up large financial institutions that may pose a risk to the financial system because of their size. The **Volcker Rule**, another key component of the Dodd-Frank Act, restricts the ways that commercial banks can invest in hedge funds, private equity funds, and proprietary trading strategies, and regulates trading in derivatives.

Each state also has a state securities commission and its own set of securities laws called **blue sky laws**, designed to protect state interests and prevent fraudulent activities within a state's borders.[3] Due to the nature of governance in the United States, the SEC shares oversight authority and enforcement with each of the 50 state securities commissions. Taken together, these interlocking laws and statutes constitute the regulatory framework that governs the U.S. securities industry.

Alternative investments fall under the SEC's regulatory authority, although historically these products have been lightly regulated. SEC regulation of alternative investment products has been premised on the assumption that these products are designed for wealthy individuals who are sophisticated investors or who, at minimum, have sufficient resources to protect themselves. When combined with the premise that government should play a limited role in overseeing private affairs—with the exception being issues of fraud or insider trading—regulation of alternative investment products within the United States is based more on monitoring and controlling systemic risks to the U.S. economy than on the protection of individual investors. A similar theme is found in hedge fund regulation as well as in private equity investment products.

34.2.2 Regulation of Hedge Funds in the United States

The regulatory scheme is governed by certain core operating principles. The first is the elimination or control of fraud, and the second is the discouragement of insider trading.[4] Both of these principles are embodied in the Securities Exchange Act and have their genesis in the U.S. stock market crash of 1929.[5] The Securities Act of 1933 and its sister statutes were enacted as a direct response to the market crash and ensuing economic turmoil, and were premised on the belief that fraud, a lack of transparency, and insider trading undermined confidence in the financial system and threatened the very social system and economic well-being of the country. In a similar vein, it was the abuses of regulatory loopholes leading to the financial crisis of 2007–8 that prompted the United States to respond with the passage of the Dodd-Frank Act in 2010, which increased regulation and disclosure requirements for investment advisers.[6]

Regulation of hedge funds in the United States begins with limiting access. Under the Securities Exchange Act of 1934 as modified by the Dodd-Frank Act, there are several important changes of which advisers need to be made aware.

- In order to invest with a hedge fund, sections 3(c)1 and 3(c)7, respectively, of the Investment Company Act of 1940 require that an individual must be an accredited investor or a qualified purchaser. An **accredited investor** is defined as a natural person who either (1) has a net worth exceeding USD 1 million, excluding the value of his primary home; or (2) has earned at least USD 200 thousand singularly in each of the past two years or USD 300 thousand jointly with a spouse. A **qualified purchaser** is defined as an individual investor with USD 5 million or an institutional investor with USD 25 million in investable assets; there is no further requirement.
- The SEC also requires those persons who file Form 13F to report certain investments.[7] This helps address systemic risk to the United States economy by compelling institutional advisers to disclose information to the public and thus increase investor confidence.
- There is increased liability for aiding and abetting: Whereas prior to the Dodd-Frank Act the SEC had to establish knowing and substantial assistance to the primary violator to convict a person of aiding or abetting, the SEC now merely has to show recklessness, a much reduced standard. In addition, the person who aids or abets a violation by another person is now liable to the same extent of fines and penalties as the primary violator.

Registration with the state or the SEC imposes substantial disclosure and regulatory requirements on a hedge fund manager. These requirements are examined further in the next section.

34.2.3 Hedge Fund Registration in the United States

As noted previously, the U.S. regulatory scheme is essentially a joint venture between the SEC and the states' securities commissions. In this section, we outline the choices and rules governing registration. It is important to keep in mind that each state commission will have its own rules and regulations, and there may be wide variances between and among the states. Thus, care should be taken in understanding and complying with each jurisdiction's regulations.

A hedge fund manager who is unable to rely on a registration exemption (or chooses not to do so) must register as an investment adviser with either a state commission agency or the SEC. The amount of assets an investment adviser has under management will determine whether the adviser must register with the SEC or one or more states. This calculation of assets is referred to as regulatory assets under management (RAUM) and differs from the calculation of assets under management (AUM) made in Form ADV 2A, which is a disclosure document made available or given to clients. **Regulatory assets under management (RAUM)** is defined as the securities portfolio (i.e., a portfolio in which at least 50% of the total value consists of securities, cash, and cash equivalents) to which the adviser provides continuous and regular supervisory or management services. The securities portfolio includes proprietary assets, assets managed for no compensation, and assets of foreign clients.

Under the Dodd-Frank Act, fund managers must use the following guidelines to determine whether and with which government agency they must register:[8]

- With the SEC:
  - Manages hedge funds whose AUM is between $25 million and $100 million and maintains a principal office and place of business in a state that does not require the registration of investment advisers
  - Manages hedge funds whose AUM is between $25 million and $100 million and maintains a principal office and place of business in a state where the investment adviser would not be subject to examination by the state securities commissioner
  - Manages only hedge funds whose AUM is greater than $100 million and maintains managed accounts
  - Manages hedge funds whose AUM is greater than $150 million and does not maintain managed accounts
- With a state commission:
  - Manages hedge funds whose AUM is between $25 million and $100 million and maintains a principal office and place of business in a state that requires the registration of investment advisers

Hedge fund managers with AUM between $25 million and $100 million are considered midsize advisers and are given a $10 million buffer so that they do not have to continually switch back and forth from state and SEC registration as their AUM fluctuates.

The United States offers an open regulatory platform to all investors who comply with its laws. Non-U.S.-based hedge funds with more than 15 U.S. clients and investors with assets under management of more than $25 million have to register with the SEC. Once a hedge fund determines whether it must register with the SEC, it must complete and file Form ADV. The form consists of two parts, along with various schedules and disclosure reporting pages. Part 1 provides information about the hedge fund, its manager, and all associated persons, and is primarily designed for use by regulators for administrative purposes. Part 2 requires information that must be provided to the fund's clients and is thus intended primarily for them (e.g., the types of advisory services provided, advisory fees charged, the fund manager's affiliations with other securities professionals, whether the fund manager effects securities transactions for clients, how securities are analyzed, whether the fund manager has brokerage or investment discretion, and potential conflicts of interest). In many cases, most of the information required in Part 2 is already provided in the fund's private placement memorandum or other offering documents.

Many hedge funds typically deliver Form ADV, which is a required submission by investment advisers. **Form ADV** specifies the investment style, assets under management, and key officers of the firm. The form must be updated annually and be made available as public record for companies managing in excess of $25 million. Form ADV is divided into two parts. The first part discloses specific information is important to regulators. The second part acts as a disclosure document for clients of the business and includes such information as services provided and fees levied, whether the investment adviser acts as a broker-dealer and transacts securities, and so on. For example, the following information is provided:

- How the adviser charges for services (hourly, fee-for-service, percentage of assets under management)
- How much the adviser charges for services (if it's hourly, what is the hourly rate?)
- What services the adviser offers (if the client is paying 1% per year, does that include comprehensive financial planning or just portfolio management, with financial planning services at an additional fee?)
- The adviser's investment philosophy (how investments are chosen and how allocations are decided)

34.2.4 Compliance Culture and Policies and Procedures in the United States

Under **rule 206(4)-7**, each registered adviser must establish an internal compliance program designed to facilitate performance of the adviser's obligations under the Dodd-Frank Act. As part of this, each advising company must have a chief compliance officer (CCO) who develops compliance procedures for the adviser. The role of the CCO is to be primarily responsible for overseeing and managing regulatory compliance issues. In order to fulfill these duties, the CCO must ensure that the fund is complying with the regulations in whichever country the fund does business. A CCO may be an employee with other duties, but it is essential that this employee be knowledgeable about the Dodd-Frank Act and have the authority to implement compliance procedures. Fostering a culture of compliance is important, as some of the most common violations that lead to SEC investigations include misrepresentation or omission of important information about securities.[9] Perhaps more important, because the SEC has the ability to sanction an adviser whose subordinate has violated a statute, adequate compliance procedures can allow an adviser to take refuge under the safe harbor provision.[10]

34.2.4.1 Duties of the Chief Compliance Officer In addition to setting the general tone for the organization about the importance of compliance in conjunction with senior management, the CCO is responsible for the following specific duties in administering the compliance program:

Compliance (or forensic) testing and reporting: Once the compliance policies and procedures are in place, the CCO is responsible for testing the effectiveness of such policies and procedures and reporting the results to senior management. This process is critical to maintaining the health of the compliance system, identifying new issues, and addressing them effectively. The general areas in which the CCO should test and document are:

- Performance management and trade allocation
- Brokerage arrangements and best execution
- Valuation
- Personal trading
- Advertisements and communications with investors

Review of all marketing materials and other documents: The SEC regulates advertising by hedge fund managers by generally prohibiting an investment adviser from engaging in fraudulent, deceptive, or manipulative activities.[11] It should be noted that this prohibition applies to both registered and unregistered investment advisers. The term *advertisement* includes any written communication addressed to more than one person, or any notice or other announcement in any publication or by radio or television, that offers any analysis, report, or publication regarding securities; any graph, chart, formula, or other device for making securities decisions; or any other investment advisory services regarding securities.[12] Electronic advertisements are also included in this definition. Some examples that fall under this broad definition are marketing presentations, monthly performance letters to investors, offering memorandums, and any document designed to maintain existing clients or solicit new ones.

In order for a hedge fund manager to refrain from violating the rule against fraudulent, deceptive, or manipulative activities, the CCO must review all documents that communicate information to investors to make sure they conform to regulatory guidelines. There are at least two important issues to consider. First, hedge fund managers are not allowed to use testimonials or endorsements from investors. Second, performance must represent long periods of time, not highlight specific time frames or only successful trades.

When fund managers include performance results in advertisements or written communications, they must also provide the following to avoid misleading investors:

- The effect of material market or economic conditions on the results portrayed (e.g., it would be misleading to report that the fund manager's equity accounts increased 10% without disclosing that the equity market was up 40% during the same period).
- Performance net of fees (e.g., advisory fees, brokerage commissions, and other expenses that a client or fund would have paid), with exceptions: (1) Gross performance figures may be presented if they are given side by side with net figures, (2) custodial fees need not be netted out, (3) performance results may deduct the highest fee charged to a client or fund as long as disclosure is made explaining how the performance was calculated, and (4) gross performance results may be presented in one-on-one presentations to wealthy individuals, pension funds, universities, and other institutions.

Record keeping: Registered investment advisers have extensive obligations with regard to record-keeping rules, and the CCO is responsible for making sure that all supervisory persons understand their obligations regarding the proper maintenance of certain books and records.[13] The CCO will also find that documenting nearly every issue and concern that arises during the course of business is critical to success in running an airtight compliance program and having peace of mind in the event of a possible SEC inspection. In addition to keeping typical business accounting records, a hedge fund manager must generally maintain certain books and records required by the SEC in light of the special fiduciary nature of the business.

Annual review: On an annual basis, an investment adviser must review policies and procedures to determine their effectiveness.[14] (A more frequent review of the compliance program is also appropriate if certain events necessitate immediate changes.) The CCO will conduct the review, which should take into account and give a detailed report of the following:

- The compliance matters that arose during the previous year, including any violations of the policies and procedures
- Any changes in the business activities of the adviser or the adviser's affiliates
- Any changes in the Investment Advisers Act or other applicable regulations that might impact the business and necessitate an amendment in the policies and procedures

34.2.4.2 Code of Ethics A registered hedge fund manager must establish, maintain, and enforce a written **code of ethics** that refers to the adviser's fiduciary duties to clients.[15] The CCO is responsible for drafting the code of ethics, having it approved by senior management, effectively communicating the code to all supervisory personnel (e.g., in an initial presentation), implementing and enforcing the code, and amending the code periodically as issues arise. The code of ethics must include, at a minimum, the following:

- Standards of business conduct for supervised persons that require them to comply with applicable federal securities laws
- Provisions that require access persons, defined as those who have access to nonpublic information, of the registered fund manager to report personal securities transactions periodically and require the fund manager to review these reports
- Provisions that require the CCO or another designated person to preapprove investments by the access persons in IPOs or limited offerings
- Provisions that require supervised persons to report violations of the code of ethics to the CCO or another designated person
- Provisions that require that each supervised person be provided with the code of ethics and acknowledge receipt of it
- Provisions that require keeping records of any violations of the code as well as any actions taken against violators

34.2.4.3 Three Types of SEC Inspections The Dodd-Frank Act amended section 204 of the Investment Advisers Act to provide for three types of inspections: regular periodic inspections, cause inspections, and sweep inspections. Regular periodic inspections are generally based on an adviser's promotional materials, including what is written in Form ADV, and are looking to ensure that there are no misleading or fraudulent statements. It is important to note that these inspections have taken on new teeth with the SEC Division of Enforcement's announcement of the creation of five departments dedicated to particular areas of securities law. These are the Asset Management Unit (AMU), the Market Abuse Unit, the Foreign Corrupt Practices Unit, the Municipal Securities and Public Pension Unit, and the Structured and New Products Unit (now the Complex Financial Instruments Unit).[16] These units are designed to more effectively monitor the various investing strategies through a better understanding of the complex products on the market and, as a result, detect violations more quickly. Once violations have been found, the SEC can bring charges in both administrative and civil suits; the end result can be large fines and/or loss of licenses or registrations for managers and their funds.[17]

The SEC categorizes registered investment managers into two general groups for purposes of inspections: high risk and low risk. Investment managers whom the SEC deems to be high risk (e.g., those with custody of client assets, those with discretionary trading, larger advisers, those with weak internal controls, or those with a history of compliance violations) can reasonably expect more frequent inspections. SEC staff aims to conduct routine examinations of higher-risk managers every three years. Lower-risk managers are not examined on a routine basis but are instead chosen randomly for inspection.[18] If a newly registered fund manager has a high-risk profile, however, SEC staff will likely plan for an inspection within 12 months of the manager's registration. It is worth noting that the SEC also looks to ensure regular updates of an adviser's promotional materials, due to concerns about an investor being misled by out-of-date information.

Routine exams are often focused in scope, concentrating on areas that have been identified as posing the greatest compliance risk, as well as areas that may pose compliance risk for the particular firm being examined. In general, the following factors contribute to a high-risk profile for a hedge fund manager:

- Large amounts of assets under management
- Questionable responses on Form ADV
- Retail client base
- Affiliated broker
- Multiple affiliates
- Fund manager or employees with disciplinary history relating to securities violations
- Performance-based fees
- Weak compliance program

The SEC will interview members of senior management and other personnel. The CCO will normally sit in on these interviews so that she can provide information to the staff as needed and also be cognizant of what areas in the fund's business are causes for concern with the SEC. Although a fund manager should expect that every document relating to his investment activities and the operation of the fund will be subject to inspection, the SEC will at some point furnish the CCO with a document request list.

Cause inspections are inspections initiated by the SEC for a particular reason, such as when it suspects that a fund manager may be violating federal securities laws. Cause inspections may arise from a client complaint, a tip, rumors of trouble, and so on. The SEC maintains a website for filing complaints at www.sec.gov, and also accepts complaints in paper form.

Sweep inspections (or theme inspections) typically occur when the SEC focuses on a number of investment advisers located in a particular geographic area or engaged in certain activities. For example, the SEC may want to gather information on how hedge funds are handling the conflict of interest that soft dollars present and thus may select a sample of hedge funds to inspect regarding this compliance issue. Sample sizes can range from hundreds of firms to fewer than 10, depending on the issue. With the increased globalization of investment advisers, the SEC and its corresponding regulators in other countries try to coordinate examinations of advisers with global operations so that inspections can occur simultaneously in advisers' offices in several countries.

34.2.4.4 Aftermath of an SEC Inspection If the fund manager and the CCO survive an SEC inspection more or less unscathed by the mountains of documentation they have had to produce, then they will conclude the on-site portion of the exam with an exit interview conducted by SEC staff. The SEC will then continue its analysis of the fund manager's documents and information received in its offices and may follow up with phone calls to the fund manager for clarification of information or to discuss issues not brought up in the exit interview. SEC examiners aim to conclude an inspection 120 days after they have finished the on-site exam.

The best result the fund manager can hope for is that the SEC, having found no problems, sends the fund manager a no-further-action letter. The next-best scenario is that the SEC finds a violation or a possible violation of the federal securities laws and sends the fund manager a deficiency letter. This letter describes the practices or activities in question and requires the fund manager to respond in 30 days of receipt of the letter, explaining what measures have been taken in response to the violation (the fund manager may also dispute the alleged violations in the response to the deficiency letter). The worst scenario for a fund manager, short of enforcement proceedings, is that the SEC finds a multitude of significant violations, at which point the fund manager should brace himself for another follow-up inspection. In this case, the SEC will send the fund manager a deficiency letter enumerating all the deficiencies found, which must be addressed immediately and certainly before the next regular inspection.

34.2.5 Various Reporting Requirements for Hedge Funds in the United States

Hedge funds are facing increasing demands by the SEC to comply with a variety of reporting, disclosure, privacy, and information-protection requirements, many of which are in addition to those imposed on registered investment advisers.

The Securities Exchange Act of 1934 requires the filing of the following disclosure reports:[19]

- **Section 13(d)** requires an adviser who beneficially owns more than 5% of a class of publicly traded equity securities to file disclosure reports within 10 days of the acquisition of those securities, identifying, among other things, the source and amount of funds used for the acquisition and the purpose of the acquisition.
- **Section 13(f)** requires a fund manager with investment discretion of $100 million or more of publicly traded equity securities to file quarterly reports disclosing

those holdings and the type of investment and voting authority exercised by the manager.

- **Section 13(g)** provides an alternative short form to the beneficial ownership reporting scheme for acquisitions by qualified institutional investors and passive investors who acquire securities in the ordinary course of their business and not for the purpose of changing or influencing controls of the issuer.
- **Section 13(h)** requires certain fund managers with investment discretion who are engaging in certain large levels of purchases and sales of national market system (NMS) securities to file with the SEC and also provide certain identifying information to broker-dealers who, in turn, must provide transaction-related information to the SEC on its request.
- **Section 16** applies to fund managers who, investing for their own accounts or for the purposes of changing or influencing control of the issuer, own more than 10% of a class of a publicly traded company's outstanding equity securities.

The Dodd-Frank Act, for the purpose of assessing the systemic risk that hedge funds pose, has given the SEC and the Commodity Futures Trading Commission (CFTC) the mandate to adopt one of the most onerous reporting obligations required of investment advisers: Form PF.[20] Registered investment advisers with at least $150 million in assets under management are required to file the new Form PF. These advisers will have to file either quarterly or annually, depending on their level of assets or, in some cases, the types of assets managed.

Large fund advisers must provide more detailed information on Form PF than must smaller advisers. Larger advisers fall into one of three categories:

1. **Large hedge fund advisers** have at least $1.5 billion in assets under management attributable to hedge funds, and must file Form PF on a quarterly basis within 60 days of the end of each fiscal quarter.
2. **Large liquidity fund advisers** have at least $1 billion in assets under management attributable to liquidity funds and registered money market funds, and must file Form PF on a quarterly basis within 15 days of the end of each fiscal quarter.
3. **Large private equity fund advisers** have at least $2 billion in assets under management attributable to private equity funds, and must file Form PF on an annual basis within 120 days of the end of each fiscal year.

Smaller fund advisers (those who are not considered large fund advisers) must file Form PF on an annual basis within 120 days of the end of each fiscal year. **Form PF** requires information on fund size, leverage, investor types and concentration, liquidity, and fund performance. Hedge fund managers must also include information regarding their investment strategy, counterparty credit risk, and use of trading and clearing mechanisms.

Hedge funds are not required to furnish investors with any particular type of information aside from the disclosures necessary to prevent misleading investors (discussed previously). Institutional investors, however, are increasingly demanding hedge fund transparency because they owe fiduciary duties to their own investors. The transparency allows for these institutional investors to conduct the necessary due diligence and monitoring of the fund's activities to ensure that the fund adheres to its investment strategies and risk parameters. In order to meet the demands of these investors, hedge funds typically provide the following types of information to their investors:

- Inspection of fund books and records
- Quarterly or periodic letters
- Annual audited financial statements

While the registrations described in this section may be time-consuming, it is important to recognize that they are a legal obligation. Moreover, the SEC has increased its pursuit of registration violators. As of 2013, the SEC is implementing what it calls the "broken window" approach. What this means is that the SEC is interested in pursuing even minor violations to demonstrate that these will not be tolerated, in order to prevent major violations from occurring.[21] In September 2014, the SEC pursued charges against some 28 officers or major shareholders for what amounted to registration violations, with total penalties of $2.6 million.[22] In reporting this news, the SEC highlighted that even if the failure to file an accurate report is inadvertent, this is no defense and the lapse would still expose the potential violator to sanctions.

34.2.6 An Overview of Private Equity Fund Regulation in the United States

Historically, the SEC treated private equity funds similarly to hedge funds, in that they were exempt from registration due to their private and limited nature. However, following the passage of the Dodd-Frank Act, these funds found themselves without the benefit of an exemption from registration. In terms of their obligations to their clients, private equity funds have faced very different levels of scrutiny due to the inherent lack of transparency involved in their transactions once ownership in a private company has been acquired.[23] Private equity fund advisers differ from hedge fund managers in that once a private equity fund gains control of a company, the advisers are exposed to far more potential conflicts of interest and temptations to take advantage of their fund than their clients may be aware of, such as the manner in which operating partners are compensated or how management fees are assessed.[24] While there is transparency leading up to the purchase of ownership in a corporation, in the aftermath of the acquisition, advisers vary in what they do and do not disclose.[25] Also potentially problematic is that at the close of these deals, fund managers collect substantial amounts in deal fees, and had done so before the Dodd-Frank Act even though they were not registered as broker-dealers.

Under the Dodd-Frank Act, private equity funds are now required to register with the SEC under guidelines similar to those of hedge funds, although with a slightly larger AUM requirement. However, as the agency is still examining its policies with respect to these funds, the adviser should be aware that requirements are subject to change. In 2014, the SEC hired industry experts and announced an effort to examine at least a quarter of the private equity funds that registered as a result of the Dodd-Frank Act.[26] Of the funds examined by the agency, the SEC found deficient policies or inadequate disclosures in more than 50% of the funds examined. As mentioned previously, these tend to be in the area of compensation, in which a lack of continued transparency could potentially influence investors to choose a fund based on misleading information. Although wealthy people may not be in need of the protections that regulation in these collective investment schemes provides, pension plans that affect millions of Americans in both the public sector and the private sector are some of the largest investors in private equity funds. As a result, SEC registration and inspection, while burdensome, is unlikely to ease in the near future.

34.3 ALTERNATIVE INVESTMENT REGULATION IN EUROPE

Each of the European sovereign states has its own scheme for regulation and compliance. Accordingly, each has its own system of enforcement and associated penalties for noncompliance. Thus, an alternative investment manager doing business in Europe must comply with the regulatory scheme of each country it is doing business in and must subject itself to the rules, regulations, social and market concerns, and judicial authority of that country. An exception to this general rule is when a manager seeks to conduct business within the European Union (EU) and its associated member states. The EU has designed and implemented a program whereby managers who are domiciled in one of its member states can provide services and market such services to other member states under a single regulatory scheme. While the basic laws are in place, the exact contours of this regulatory scheme are still in development and represent challenges to managers seeking to conduct cross-border business in the EU, as well as to EU member states.

34.3.1 A Brief Historical View of the European Union

In its current form, the EU has been in existence since 1993, with the adoption of the Maastricht Treaty by the then member states. The stated goal of the Maastricht Treaty is to create a single economic community that incorporates social and political cooperation. It is important to understand that the EU is essentially a federation whereby different member states have agreed to varying levels of participation in the EU structure without relinquishing their basic sovereignty. By example, 19 of the current 28 member states have adopted the euro and compose the eurozone, whereas a number of other EU participants, such as the United Kingdom and Denmark, do not use the euro as their currency. The federation model of the EU also provides that member states retain all powers not explicitly given to the union through treaty. The member states of the EU have agreed to three forms of binding union legislation: (1) legislation that supersedes each member state's legislation, (2) directives that constitute a framework of objectives that member states must adhere to when creating their own laws (the member state's laws must adhere to stated principles as well as aims and not necessarily the letter of the directive), and (3) legislation relating to singular issues or concerns.

Finally, given that this vast union is relatively new, both the EU and its member states are testing the boundaries of coexistence, cooperation, and interpretation. Against this background, the EU's regulatory scheme—much like that of the United States—focuses on transparency and systemic risk to the EU financial system. More important, it is an attempt to promote the rule of law and thus dispute resolution through integration of the financial regulatory bodies of sovereign states into a cohesive whole that has common rules and principles.

34.3.2 An Overview of Regulations for Alternative Asset Managers in Europe

European regulation of alternative investments can be complex. First, as noted in the previous section, not all European nations are members of the EU and not all EU members have the same status within the EU. It is worth reiterating that the EU is a federation of sovereign states that have, through various treaties, relinquished certain elements of their sovereignty. Through these treaties, the EU has sought to create a federation that reflects a cohesive whole with common goals, principles, and core value systems. As those treaties and core values are related to the EU's financial markets, the stated goal is to create a democratic free market environment that is based on the rule of law and as such promotes transparency, integration, predictability of dispute resolution, and fundamental market fairness. When combined, these elements provide for meaningful oversight and the regulation of systemic risk for both the EU and its member states.

Within the terms of the relevant treaties, the rules and regulations relating to alternative investments for distribution across the EU were passed by the European Parliament and promulgated by the Financial Conduct Authority (FCA).

In turn, the FCA proposed the Alternative Investment Fund Managers Directive (AIFMD), which was adopted and ratified by participating states between 2011 and 2013. Compliance with the AIFMD is intended to provide investors comfort as to the market integrity of financial products purchased from such firms, and to provide investment managers with common regulations across EU member states, as well as a "passport" to market and deliver their services across EU borders. In so doing, the model significantly reduces the administrative and operating costs of investment managers conducting business in multiple jurisdictions.

The **Alternative Investment Fund Managers Directive (AIFMD)** reflects the basic regulatory model of a directive and, in so doing, provides for each member state to carry out the following:

- Adopt and implement the aims of AIFMD
- Provide regulatory supervision and enforcement over alternative managers within its jurisdiction
- Certify that managers within its jurisdiction who desire an AIFMD passport are compliant
- Work within AIFMD guidelines in resolving enforcement and oversight disputes with member states as well as the EU

Within this model, the member state where the manager is domiciled has initial authority over all compliance and enforcement matters. However, the model contemplates that both regulatory oversight and enforcement are iterative processes and can become a matter of negotiation between the home state and relevant EU institutions, including in some instances other member states. Thus, the model provides a mechanism for dispute resolution among these political entities.

Also, in many ways the AIFMD model is a significant departure from traditional regulatory and disclosure models in the EU. The model is a reaction to the global financial crisis of 2007–8 and seeks to monitor and control a number of the activities that contributed to that crisis. There is a strong emphasis on (1) valuation,

(2) counterparty risk, (3) operational risk, and (4) liquidity. Furthermore, in a complete departure from past regulatory schemes, the model provides oversight for issues such as character and reputation, and calls for a written remuneration policy for all employees. Some of the essential aspects of this developing regulatory scheme are outlined in the following sections.

34.3.3 Registration and Exemptions in European Regulation

In a review of the requirements concerning registration and exemptions, it is clear that the EU and its member states are focused on the type of firms or institutions that may be considered speculators and that are large enough to cause significant market risk. In addition, the registration requirements suggest that the model seeks as much transparency as possible regarding the operations of fund managers to ensure that both regulators and investors have sufficient information to quantify operational, market, and counterparty risk. The lack of proper analysis of operational and counterparty risk was a major contributor to the 2007–8 financial crisis.

34.3.3.1 Registration in European Regulation As noted, the AIFMD provides a template for the minimum requirements for conducting business among the member states with a passport.[27]

Managers and firms that invest in publicly traded stocks or shares fall under an EU regulation scheme known as the **Undertakings for Collective Investment in Transferable Securities (UCITS)**, which refers to a series of EU directives that established a uniform regulatory regime for the creation, management, and marketing of collective investment vehicles in the countries of the EU. UCITS funds typically invest in securities listed on public stock exchanges and regulated markets. The UCITS directives were designed to bring European investors a wide offering of funds together with a high level of investor protection. Investors can invest in any UCITS fund that has been registered for sale. Before the first UCITS directive, most investors were largely limited to funds offered by fund companies based in their country of residence. The AIFMD applies to non-UCITS investments such as private equity, venture capital, hedge funds, and real estate funds.[28]

It is worth noting that the AIFMD is meant to be a companion directive to the UCITS program. UCITS products have approximately €7 trillion under management. The program was initially launched in 1985 and has been subject to a number of amendments and restatements; in 2014, the EU adopted UCITS V. The overall goal of the UCITS program is to facilitate cross-border distribution, on a "passport" basis, of investment products, relying on the regulatory oversight of a single member state. As suggested, UCITS is a very complex and evolutionary program. Also, as a directive, such legislation carries a host of associated conflicts. Member states have taken full advantage of this structure to protect their home-based managers, thus limiting the effectiveness of the program in creating a uniform regulatory framework of certain investment products.

One of the key elements of the UCITS structure is the due diligence requirements on service providers, which attempts to standardize operational due diligence by creating a transparent platform subject to independent verification. Another is the requirement that all of a manager's positions be marked to market on a daily basis, thus providing both daily valuations and daily liquidity to investors. The UCITS program has been used by Japan, South Korea, and some EU countries as a method of evaluating bank holdings and determining into which tier to place certain investments when accessing a bank's viability. Also, the EU envisioned UCITS as a method to expand its influence and distribution into Asia and South America and has pushed these geopolitical zones to adopt UCITS standards.

It was thought that UCITS could be an effective vehicle for the distribution of alternative products, such as hedge funds and private equity. A number of firms have developed platforms to accomplish this end. However, it has been found that the strength of the UCITS program actually works against using this vehicle for alternatives. The requirements of daily marks and daily liquidity have shown that only equity long/short funds and some public debt funds have been able to gain meaningful distribution traction. Funds that are based on assets that are difficult to price on a daily basis have attempted to use swaps and derivatives as a means of fitting into this regulatory scheme. However, most management platforms have found that the tracking error between a manager's flagship fund and the results of this approach is so significant that it yields a completely different risk-reward calculation and is in effect a new product, making the historical performance of a manager's flagship product a meaningless comparison. Against this background, the EU, its member states, and investment companies have pushed to develop the AIFMD standards.

34.3.3.2 Exemptions to Compliance in European Regulation The AIFMD provides a number of exemptions to compliance.

Family offices: The AIFMD model does not view family offices that do not solicit outside capital as alternative investment funds (AIFs); thus, they are exempt from AIFMD compliance.[29]

Non-marketed funds: These regulations do not apply to holding companies, pension funds, employee participation and savings funds, and other institutions that manage funds yet do not market such funds to investors as alternative investments.[30]

Minimum thresholds: Alternative asset management firms whose assets under management are below certain thresholds do not have to obtain full AIFMD authorization. Such funds are considered "sub-threshold" and qualify under the exemption of the "light regime"; as such, managers must register these funds with their local authorities. This exemption applies to funds that manage less than €100 million with leverage or less than €500 million without any leverage. Both of these types of funds are required to have no redemption rights for five years.[31] The disadvantage to light regime funds not having a full AIFMD registration is that they will not be granted a passport and are thus subject to direct oversight by each member state where they are doing business.

34.3.3.3 Obtaining Registration in European Regulation Alternative investment fund managers (AIFMs) must apply for compliance authorization to the appropriate authorities in their home state. AIFMs must provide the necessary information required by Article 7 of the AIFMD in their application.[32] The application requires information about those conducting the business of the AIFM; the shareholders/members; organizational structure; remuneration policies and practices; delegation and sub-delegation arrangements; investment strategies; types of funds; types of assets; use of leverage; risk profile; the location of the master alternative investment fund (AIF), as well as other states where AIFs shall be established; the investment management agreements of each AIF; depositary arrangements; and any other information that is material to the management of an AIF.[33]

34.3.3.4 AIFMD Fund Directives and Fund Requirements A significant subtext of the AIFMD and its disclosure requirements is to provide investors and regulators with tools to properly gauge the health and viability of a fund manager as an ongoing business concern through the introduction of capital requirements. This is a regulatory requirement that does not exist outside of the commercial banking industry.

- *Initial capital:* Internally managed funds must have initial capital of at least €300,000, while externally managed funds must have initial capital of at least €125,000.
- *Insurance:* All AIFMs must have professional indemnity insurance against liability arising from professional negligence, at an amount that is appropriate to the risks covered. Required coverage is 0.7% for any single claim and 0.9% of the value of the portfolio's prior 12 months.
- *Liquidity thresholds to meet fund commitments:* The AIFMD requires AIFMs to monitor the liquidity of the portfolio. AIFMs must also document their liquidity policies and the methods by which they assess the risks that could create material risk to the liquidity of the fund.[34] The directive also calls for AIFMs to periodically conduct stress tests to assess liquidity exposures within their portfolios.[35]
- *Limitations on leverage to manage risk:* In its Level II Directive, the European Securities and Markets Authority provides detailed methods for funds to use when calculating leverage. AIFMD has assigned limitations on leverage to be set by the competent authorities in each home state.[36] The amount of leverage that a fund may use will vary according to the type of asset that a fund manages, as risk varies by asset class.
- *Reporting of profits according to the AIFMD's definition of calculating profits:* The AIFMD framework provides rather complex guidelines for the reporting of profits.
- *Valuation:* The AIFMD imposes valuation methods and requirements as well as the use of third parties to minimize fraud. Valuation reporting is also intended to avoid AIFMs earning fees based on inflated valuations of assets under management.
- *Management:* Regulations require a minimum of two senior managers who are sufficiently experienced in managing money and are of "good repute."
- *Remuneration:* The AIFMD requires written guidelines regarding remuneration policies for senior management, as well as employees whose roles and responsibilities have material effects on the risk and return of the funds. The policies also include employees whose compensation is of a similar level to senior management.

34.3.3.5 Marketing Materials in European Regulation The disclosure requirements for offering memorandums are quite extensive and in many ways mirror those of the U.S. Securities and Exchange Commission. In addition to mandating the disclosure of depositaries, auditors, and service providers, the model specifies the following requirements for offering memorandums:

Overall Information

- There must be a description of the investment, including strategy, objectives, and restrictions; how units or shares are to be sold; and all fees to be charged to investors.
- Existing funds must include the most recent annual report, the historical performance of the fund, and the most recent net asset value of the fund (including the market price of the units or shares).

Fund Management

- There must be a description of how the manager may change investment policies and procedures
- Methods of valuation, including the procedures for valuing hard-to-value assets such as derivatives and privately held companies, must be described. If the fund permits leverage, there must be a description of the leveraged items, the risks associated with that leverage, and the collateral that the fund is permitted to use for that leverage. How liquidity will be managed and the investor's redemption rights must also be addressed. A depositary will be appointed to independently monitor cash flow, be responsible for safekeeping of financial instruments, and make sure that assets reported are, in fact, owned by the fund.

Compliance

- AIFMD compliance disclosure: Legal implications of all contractual arrangements must be described, along with identification of depositary, auditor, and other material service providers and identification of and material arrangements with the prime broker. If the manager plans to delegate any management of the fund, such plans must be described. There must be a description of how the fund will assure the fair and equal treatment of all investors.

34.3.3.6 Formal Risk Management and Accountability in European Regulation

The AIFMD provides guidelines for determining levels of transparency and safeguards regarding risk. In recognition of the fact that risk management is one of the key tools to effective oversight, the EU and its member states created the **European Systemic Risk Board (ESRB)**, which is an additional European regulatory body to oversee risk. AIFMD compliance regarding risk will fall under the ESRB in the EU and the FCA in the UK.

The ESRB is responsible for the macro-prudential oversight of the financial system within the EU in order to contribute to the prevention or mitigation of systemic risks to financial stability. The term *macro-prudential regulation* characterizes the approach to financial regulation aimed at mitigating the risk of the financial system as a whole. For this purpose, the ESRB is expected to carry out certain tasks, including (1) collecting and analyzing all the relevant and necessary information, (2) identifying sources of systemic risk, (3) issuing warnings about where such systemic risks are deemed to be significant, and (4) issuing recommendations for remedial action in response to the risks identified.

The need for such stringent risk management is intended to ensure that AIFs are appropriately managed in order to protect investors. Further, it enables regulators to test whether the fund is being managed with prudence and care, whether liquidity mandates are being maintained, and whether the AIFM is within compliance at all times.

The rules regarding risk management include assessment of risks based on the amount of leverage used. Rather than provide specific guidelines for the amount of leverage to be used, the AIFMD requires "reasonable" limits. Accordingly, the ESRB on a case-by-case basis has the ability to establish guidelines without being hindered by a preexisting standard.

The guidelines acknowledge that risk management is both a quantitative and a qualitative exercise, and that risk limits evolve depending on the complexity of the portfolio and external market, asset class, or political conditions. Thus, risk monitoring is expected to be ongoing and includes periodic stress tests, backtests, and random scenario analysis for potential variances in market conditions for each AIF and AIFM. The guidelines further provide that risk policy must take into account the measurements and analysis of risk relative to market risk, liquidity, counterparty risk, and other risks, including operational risk. Operational risk involves such items as network security, monitoring of depositaries, prime brokers, and other risks that are not directly financial, including cyber security.[37]

At the individual manager level, the guidelines provide that risk management must be independent and separate within the AIF to avoid conflicts of interests, and must be separated from operating units and portfolio management.[38] In order to make sure that those responsible for risk management remain independent, they should not be supervised by a partner or an employee who is responsible for the performance of the operating units or who has a role in portfolio management.[39] Those responsible for risk management functions cannot be engaged in performance activities of operating units and/or functions of portfolio management.[40] Regarding compensation, the risk management employees' remuneration must be linked to the meeting of the objectives of the risk management platform and not to the returns of the fund. The goal of these provisions is to ensure that risk management is not co-opted by the fund's portfolio managers. Further, risk management remuneration is to be overseen by the remuneration committee if one exists within the AIF.[41]

34.3.3.7 Required Reporting in European Regulation The reporting requirements under the AIFMD represent another tool by which the regulations are intended to promote transparency. The directive regarding annual reporting in Chapter IV, Article 22, of the AIFMD is named "Transparency Requirements."[42] Reporting under the AIFMD consists of annual reports to investors and regulators, periodic reports based on the size of the AIF/AIFM, and reports that must be filed when there is a material change within the AIF.

Annual report: Each AIFM is required to provide an annual report no later than six months after the end of the financial year, which must include a balance sheet and an income statement. In addition, AIFMs must include material changes for the past year and details regarding remuneration.

Reporting to authorities: The AIFMD requires reporting of certain items to competent authorities, which will enable those authorities to effectively monitor AIFMs. These reports will vary in terms of size, leverage, liquidity, and what is specifically requested by regulators.

Because there was no centralized platform for either valuing these instruments or monitoring their deployment within portfolios, derivatives played a significant role in the 2008 financial crisis. As a result, authorities want to see a reporting mechanism that reinforces the AIFMD's commitment to transparency and the avoidance of systemic risks.

The AIFMD requires that AIFMs disclose the instruments traded by the AIFs, as well as the markets in which these trades are made. The AIFMD requires a certain level of checks and balances to both avoid conflicts of interest and increase transparency of the fund's holdings. Independent compliance functions as an effective means to avoid fraud by AIFs and AIFMs. The AIFMD directive regarding depositaries is very much the same as the UCITS directive regarding depositaries. Additionally, regulators want to assess the concentrations of assets of all the AIFs that an AIFM manages. These required disclosures are to include items that have an effect on liquidity, the results of stress tests, leverage, and multiple types of risk, such as market and counterparty risk.

Frequency of reports to authorities: The required frequency of reporting to competent authorities is determined by the size of the AIF's assets under management. Any one AIF managing over €500 million and any AIFMs managing AIFs that manage over €1 billion must report quarterly. AIFMs with AIFs that manage less than €1 billion are required to report on a semiannual basis.[43]

Ongoing reporting to authorities: In order to minimize and oversee risk, the authorities expect reporting when certain changes occur that may cause material changes to an AIF. Article 24 of the AIFMD outlines most of the matters important to these regulators. The most important material risk disclosures are leverage, liquidity, and material changes in values of assets that may have an adverse effect on the fund.[44]

Additional reporting for private equity funds: Private equity funds or any funds that invest directly in private companies are required to report certain actions regarding their ownership. Articles 26 through 30 call for AIFMs to report whenever one or more of their AIFs acquire more than 50%, or control, of a nonlisted company. In addition to reporting the control of ownership, the manager must also report how the acquisition was made, if leverage was used, where the leverage came from, the intentions for the future of the company, effects on employment, and employment conditions.[45]

AIFs that acquire nonlisted companies are subject to asset stripping rules. **Asset stripping rules** prevent an AIF from making a controlling private equity investment, having the nonlisted company take a loan, and then distributing the loan proceeds to themselves, and thereby creating leverage that may or may not create unnecessary risk.[46] Asset stripping rules apply for a period of two years after acquiring control of a nonlisted company, during which time the AIF and/or AIFM is not permitted to influence or set distributions, reductions in capital, or share redemptions. These rules also seek to ensure that managers do not unjustly enrich themselves at the expense of the fund and its investors.

34.3.4 Enforcement of European Regulation

The AIFMD model recognizes that each state has its own culture and administrative/judicial procedures and associated precedents. It also acknowledges that in order to be successful in its goals, there needs to be a level of standardization.

34.3.4.1 The Role of the European Securities and Markets Authority The **European Securities and Markets Authority (ESMA)** was created by the EU parliament with the power to write technical standards and bring about systems of mutual recognition. ESMA's role in the AIFMD is one of legislation.[47] The enforcement of these rules is left to the member states; however, should member states not enforce the AIFMD, ESMA may bring them before the EU to enforce the EU's intent of equalizing the playing field regarding alternative investment regulations. All member states have a national regulator who sits on ESMA's board. To the extent that two member states disagree on violations or enforcement procedures, the model provides for ESMA to mediate and, if necessary, take steps to bring action before the European Parliament for resolution.

The AIFMD mandates that the supervision of an AIFM is the responsibility of the "competent authorities" where the AIF or AIFM is located.[48] These competent authorities are selected by the home member state and are required to set up proper methods to monitor and enforce compliance of AIFMs and AIFs.[49] The competent authorities may investigate directly or in collaboration with other authorities, delegate relevant tasks, and apply to judicial authorities if they choose to.[50] In order to carry out such enforcement actions, competent authorities have the power to access all information from the AIF or AIFMs. Information includes phone data, mail, email, data, accounting information, and depositary information. The search or request for data includes requesting data from depositaries and accountants and anybody involved with the AIF. Host competent authorities can inspect AIFs without notice. The competent authorities can also issue cease and desist orders, freeze assets of an AIF or AIFM, request the prohibition of professional activity, suspend issues of units or shares, withdraw the authorization of an AIFM, and refer matters for criminal prosecution.[51]

However, if an AIF violates regulations or commits a crime, such as fraud, in a host state where it markets or manages a fund, the AIFMD does provide the structure for investigation and sanction. Section 4, Article 45, states that should a violation by an AIF in a host country occur, the host country is required to have the violating AIFM end the breach and inform the home member state of its actions.[52]

Should the AIF refuse to cooperate with the host country's competent authorities, the host state must inform the home member state's competent authorities as soon as is possible. Once the home member state is informed, it must take appropriate measures.[53] "Appropriate measures" are defined as providing the host state with all information requested by its authorities and adhering to any cease and desist orders. Should the AIF continue to refuse to cooperate with the host state—as well as the home state—the host member state authorities can require the AIFM to "cease managing the AIFs" and take any measures necessary to protect the investors of that AIF.[54]

In cases in which the host state and home state disagree on such issues, either party can bring the issue to the ESMA. The ESMA will facilitate negotiations between the parties and does have the power to impose binding mediation.[55] When a host state believes a violation has occurred and the violation is not within the host state's jurisdiction, the host state may refer this information to the authorities in the home state of the AIFM.[56]

An important exception to the enforcement powers of the ESMA is the sovereignty exception. The AIFMD provides that member states may refuse to cooperate if "cooperating adversely affects the sovereignty, security, or public order of the member state addressed."[57]

34.3.4.2 Penalties in European Regulation The current EU regulatory scheme does not provide for a universal set of penalties. Penalties are established within the legal framework of each member state. The AIFMD provides that the penalties should be "effective, proportionate, and dissuasive."[58] As noted previously, this regulatory scheme is evolving, and some member states have been more aggressive than others in establishing a comprehensive regulatory scheme. Thus, an AIFM should check with each member state as to the penalties before marketing or setting up an AIF in that state.

Competent authorities have the right to publicly disclose any penalties or violations unless doing so would jeopardize financial markets, be detrimental to investors, or cause what the directives refer to as "disproportionate damage to parties involved."[59] ESMA has this power and publishes annual reports on penalties and violations.

AIFMs have the right to appeal decisions of competent authorities and member states. Appeal rights are available to an AIFM if a competent authority refuses to grant authorization or takes measures against an AIFM, or if a member state imposes legal action against an AIFM.[60]

Included in Article 49(2) of the directive, AIFMs also have the right to appeal if they submit a request for authorization.[61]

34.3.4.3 The Enforcement Role of ESMA As an aside, member states are required to pass laws enacting the AIFMD. Member states that refuse are subject to the jurisdiction of the European Commission for infringement actions on the grounds of noncompliance. In the event of a dispute between a member state and the commission, either party may take the action to the Court of Justice of the European Union for resolution of the conflict.

While the AIFMD contemplates that each member state will enact its own enabling legislation, it also provides that if member states do not enforce the substance of the AIFMD, ESMA may use its enforcement powers to ensure that the AIFMD is properly enforced by its member states. To this end, ESMA has the power to investigate an EU regulator.

When a fund violation is suspected, the process differs based on whether it occurs within or outside the AIFM domicile. If the violation occurs within the AIFM domicile, the first step requires the local member state's governing body to initiate an investigation. For those violations that occur outside the AIFM domicile (e.g., a UK-domiciled fund), the FCA would cooperate and even conduct a joint inspection with the host member authority. This model is designed to protect all investors, regardless of their country of residence.

Overall, ESMA is designed to monitor and enforce a number of subtexts within the AIFMD model. First, the disclosure requirements acknowledge that for many alternative investments, there can be a sustained period of time in which investors are only receiving accounting returns, and such returns may not correlate with actual monetary returns. Next, it provides a mechanism to achieve transparency within the operational and administrative platforms of both member state regulatory bodies and individual AIFMs or AIFs. Finally, it is a mechanism for standardizing the rule of law between and among member state regulatory bodies.

34.3.5 Non-EU Managers in Europe

Currently there is a framework for a non-EU consideration model in which the AIF is active. The large alternative managers—such as Blackstone, TPG, BlackRock, and Apollo—are domiciled in London and thus have applied for or have AIFMD status.

Non-EU AIFMs who have worked in the EU before the passage of AIFMD are permitted to continue to do so. From 2013 until 2018, member states have the discretion to allow private placement memorandums with a certain amount of AIFMD compliance.[62] AIFs and AIFMs must be from countries with approved status with ESMA; for example, the SEC, CFTC, or Cayman Islands Monetary Authority have signed off in order for an EU regulator to follow that AIF back to its country of origin for reasons such as investigation or prosecution. Many other countries have signed treaties to this effect. Non-EU AIFMs must find out what the requirements are in each member state in which it plans to market or manage funds. However, non-EU AIFs must still adhere to basic AIFMD compliance. These items include reporting, disclosure, transparency, and—in the case of private equity firms—reporting on asset stripping.

As noted at the beginning of this chapter, regulations are a compromise between and among governments and business interests as well as social and monetary policies. Correspondingly, regulatory oversight and enforcement are the way governments and political entities ensure adherence to these compromises and policies. Because they exist in a political environment of compromise, regulations are fluid and subject to change through both interpretation and implementation. The AIFMD clearly shows the tensions and challenges associated with oversight and enforcement in the European community. As a consequence, many countries have been slow to act, and money managers have elected to stay on the sidelines until there is more clarity. The possibility of give-and-take within the enforcement system, coupled with the uncertainty regarding penalties and due process, actually serves to threaten one of the core governing principles of the EU: the rule of law, and thus the predictability of dispute resolution. Undoubtedly these issues will be resolved, as there are too many interests that need a clear resolution. For the moment, the regulatory scheme is challenging to the member states, the EU, investors, and money managers.

34.4 HEDGE FUND REGULATION IN ASIA

A number of Asian governments have recently become more desirous of attracting hedge funds to their shores and have been designing regulatory regimes accordingly. Hong Kong and Singapore require hedge funds to register with the appropriate government regulatory body, but the costs of starting and running a hedge fund in these two countries are much lower than in New York (as are the income tax rates). Where Singapore is more innovative and proactive in its oversight, however, Hong Kong is more conservative and wields tighter regulations. Hong Kong and Singapore, in particular, seem to want to become the next centers for hedge funds, to rival those of New York and London.

Asia is slowly opening up to the idea of foreign investments. South Korea has recently allowed onshore hedge funds for the first time and expanded its regulation of offshore funds wishing to do business in the country. By contrast, Taiwan does not allow investments by offshore hedge funds, and Shanghai is largely off-limits to investors outside mainland China. As of August 2014, Singapore, Malaysia, and Thailand have launched the ASEAN CIS Framework to facilitate cross-border offerings of collective investment schemes to retail investors. However, this framework still requires a fund to comply with regulations in its host county or, if it is foreign, that the foreign country has in place similar legal protections for investors. As a result of the wide variance in regulations and licensing requirements, Western hedge funds will likely find procuring investments from Asia to be far more challenging than the process is in Europe, because there is no common regulatory framework.

34.4.1 Hong Kong

In Hong Kong, the **Securities and Futures Commission (SFC)** is the primary regulator of hedge funds.[63] The SFC is primarily concerned with transparency in investments, and many of its regulations are geared toward ensuring that investors have adequate information to make informed decisions.[64] Under the Securities and Futures Ordinance (SFO), a hedge fund manager may operate in Hong Kong once it registers with the SFC, provided the manager meets other licensing requirements.[65] In addition to obtaining licenses for the firm and individuals, hedge funds must follow these requirements:

- *Capital requirements*. The fund manager must meet capital requirements and pass a licensing examination. Under the SFO, a hedge fund licensed for asset management is generally expected to have at least USD 100 million in AUM, which can include proprietary funds.
- *Relevant experience*. The management company must have the requisite competence and experience, as well as be suitably staffed to properly manage the risks and operational issues in connection with its hedge fund business. There must also be at least two key personnel, each with five years of relevant experience and at least two years of specific experience in the fund's strategy. Recent guidelines have found that general experience acquired through academic research or marketing would likely be insufficient to satisfy this requirement.
- *Authorization for public offerings*. Hedge funds must be authorized by the SFC before offering or inviting the public to acquire shares or interests in them.[66] Alternatively, a hedge fund can seek to qualify for an exemption.
- *Adequate controls*. Hedge funds must demonstrate that they have internal controls in place that will compensate for that company's business risk profile. They must also demonstrate that their representatives and agents have sufficient knowledge and experience in dealing with hedge funds.

34.4.2 Singapore

The **Monetary Authority of Singapore** (**MAS**) is the primary hedge fund regulatory agency in Singapore. The stated goal of the agency is sustainable growth; therefore, many of the regulations are geared toward risk management.[67] To that end, advisers who wish to operate funds in the country are required to be licensed with the MAS or an accredited investor within the framework of ASEAN CIS. In addition to licensing of the firm, hedge funds must meet the following requirements:[68]

- *Relevant experience.* Managers should have experience in the relevant investment scheme, and at least two executives should have at least five years of experience in hedge funds, of which three years should be in the management of funds of funds.
- *Capital requirements.* Singapore has minimum subscription requirements, meaning that a single hedge fund should be offered with a minimum subscription of SGD 100,000 per participant.
- *Adequate controls.* Managers of these funds must demonstrate proper internal controls that will compensate for business risk. These controls include ongoing monitoring of the fund (including regular audits), an adequate due diligence process for the selection of underlying funds, and annual certification to the MAS that its procedures are set out in its prospectus.
- *Limitations on investment.* A single hedge fund is allowed to invest only in another single hedge fund that is not part of a feeder scheme. A fund of hedge funds is allowed to invest in an underlying fund of funds only if that second scheme invests in other single hedge funds.

34.4.3 South Korea

In South Korea, the regulatory body is the **Financial Supervisory Service** (**FSS**), which operates under the Financial Services Commission (FSC). It is responsible for inspection of financial institutions as well as enforcement of relevant regulations as directed by the FSC. Local hedge funds are quite new in South Korea, as the previous regulatory framework regulated them out of existence. Current regulations are no less strict. In order to limit systemic risk, they provide for numerous restrictions, such as an upper limit on leveraged transactions, stringent reporting requirements, and limits on some of the strategies of these funds that are otherwise available in the United States and Europe. Other requirements include the following:

- *Licensing requirements.* Specific licenses are available for hedge fund businesses, and preexisting licensed entities (such as asset management companies that have different AUM requirements) have limited accessibility to these licenses, depending on their satisfying certain threshold requirements.
- *Relevant experience.* Certain licensed entities—such as securities firms, investment advisers, asset management companies, and foreign hedge fund managers—are eligible to apply for the license.
- *Capital requirements.* Under the current regulatory regime, foreign hedge funds are able to take advantage of more favorable regulations than their local counterparts. For example, foreign fund managers must have KRW 1 trillion in AUM

to obtain a hedge fund manager license, whereas local managers must have KRW 10 trillion in AUM.

- *Adequate controls*. Hedge funds are required to set guidelines for internal controls and have specially created risk management divisions within the company.

34.4.4 Japan

Hedge funds in Japan fall under the **Securities and Exchange Surveillance Commission (SESC)**.[69] Managers are required to be licensed and register as investment managers as well as provide continuing reports on the actual business condition of the fund. Offshore funds are able to decide whether to register but may not directly market to Japanese investors; however, by law, Japanese institutions are able to invest only in Japan-registered investment vehicles, meaning that offshore hedge funds are off-limits. Additionally, Japanese regulators may, from time to time, ban the acquisition of certain types of investments. For example, due to economic turmoil, the Japanese government once imposed a ban on short-selling securities as well as increased disclosure requirements on certain types of transactions.[70] While the SESC does not have the authority to initiate legal proceedings for violations on its own, it does provide for the agency to conduct so-called viable inspections, which are either with or without notice, to ensure compliance with the regulatory scheme.[71] Additional requirements are:

- *Disclosure statements*. Hedge funds must be authorized to offer securities to the public before the distribution of marketing statements. The new laws require managers to submit securities registration statements and other disclosure documents to provide investors with the information necessary to make an informed decision in the market. As a result of these rules, the SESC will inspect these disclosure statements to ensure accuracy and to deter violations.
- *Capital requirements*. The minimum capital requirement of funds is JPY 50 million for licensed discretionary investment funds. The minimum is lowered to JPY 10 million for offers made to only qualified institutional investors.
- *Adequate controls*. Hedge funds are required to establish internal legal compliance and control policies. These include adequate risk management policies and arrangements for ensuring the adequacy of responses to the results of internal audits.

34.4.5 Summary of Asian Hedge Fund Regulation

As discussed, Asia has no comprehensive regulatory scheme for alternative investments. The major jurisdictions—Hong Kong, Singapore, South Korea, and Japan—have begun to aggressively address this issue as they seek to grow and expand their home-based alternative investment industry. It can be anticipated that these regulatory models will mirror their Western counterparts and become more aggressive as they attempt to bring Western investors to their jurisdictions. In mirroring their Western counterparts and creating a viable business environment, these countries will need to foster the fundamental principles of regulation in a free market economy: transparency, fundamental fairness, and adherence to the rule of law.

34.5 CONCLUSION

This chapter examined the financial regulatory and compliance systems of the United States, Europe, and Asia. Rather than focusing on specific rules and regulations that affect participants in the alternative investment industry of these geographical areas, the chapter attempted to emphasize some broad principles that inform the regulatory scheme of these markets. The guiding principle was that regulations are a compromise between and among governments and business interests as well as social and monetary policies.

The chapter explained that beyond the individual dynamics of a given political jurisdiction, capital markets are built on trust. Financial regulations are based on principles of (1) transparency, (2) market integrity or fundamental fairness, and (3) government protection of economic and social systems through the rule of law. Although regulations will change from country to country, these guiding principles will provide the investor or industry professional with the beginning point of analysis in most situations.

Given the nature of regulation and that fact that it is subject to sudden changes, this chapter reminded readers that it is important to note that the material presented here is likely to be dated and all regulations are subject to interpretation, and thus work must be done to ensure that one understands how and why regulators are interpreting regulatory provisions at the moment. In addition, there are no global rules, so one must pay very close attention to and observe the rules of the applicable jurisdiction, never assuming to know the rules but always checking and verifying; and finally, most financial regulatory schemes provide for civil, monetary, and/or criminal penalties against individuals and institutions for noncompliance or violations.

NOTES

1. The area of compliance is a constantly changing arena, as the various regulatory agencies discussed in the chapter respond to new market conditions. As a result, the information herein should be treated as a guideline. If you have questions about the regulations in your jurisdiction, note that there is an entire industry dedicated to compliance that can help you ensure that your business is operating legally.
2. For more information about the SEC, see www.sec.gov/about/whatwedo.shtml.
3. See *Hall v. Geiger-Jones Company*, 242 U.S. 539 (1917).
4. 15 U.S.C. §§78j-l.
5. The stock market crash of 1929 occurred when speculators fraudulently pumped up the prices of stocks sold to the public and then sold their own stakes for windfall profits to the detriment of individual investors, who suffered extreme losses. The stock market crash, coupled with a massive drought in the Midwestern part of the United States, ushered in the Great Depression and one of the most stressful economic periods in American history.
6. The Dodd-Frank Act evidenced a new concern with protecting even sophisticated investors by providing greater transparency to enable investors to make more informed decisions. The preamble to the statute states that it is intended “to promote the financial stability of the United States by improving accountability and transparency in the financial system, to end ‘too big to fail,’ to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and for other purposes.” H.R. 4173.
7. Form 13F filers are institutional investment managers, such as banks, insurance companies, and pensions. 17 C.F.R. §240 13f-1.

8. Section 410 of the Dodd-Frank Act.
9. SEC.gov, www.sec.gov/News/Article/Detail/Article/1356125787012.
10. 15 U.S.C. §78o-(b)(4)(E). Failure to have in place adequate compliance procedures can lead to fines and sanctions. See, for example, SEC press release, "SEC Sanctions Three Firms under Compliance Program Initiative," October 13, 2013, www.sec.gov/News/PressRelease/Detail/PressRelease/1370540008287.
11. 17 C.F.R. 275.206(4)-1.
12. Ibid.
13. Failure to accurately maintain these records can lead to fines and other sanctions. For example, in October 2014, following the discovery of an intentionally altered trading review document filed by a Wells Fargo compliance officer, the SEC fined Wells Fargo $5 million and pursued other charges against the employee. SEC press release, "SEC Announces Enforcement Action against Former Wells Fargo Advisors Compliance Officer for Altering Document," October 15, 2014. In July 2014, the SEC brought charges against both the CEO and the CFO of a Florida company for misrepresenting their internal controls over financial reporting to auditors in order to borrow more money from creditors. SEC press release, "SEC Charges Company CEO and Former CFO with Hiding Internal Controls Deficiencies and Violating Sarbanes-Oxley Requirements," July 30, 2014, www.sec.gov/News/PressRelease/Detail/PressRelease/1370542561150.
14. The SEC has provided a summary of some of the questions a compliance officer should ask during the annual review, including risk assessment and quality control questions. SEC, "Questions Advisers Should Ask While Establishing or Reviewing Their Compliance Programs," May 2006, www.sec.gov/info/cco/adviser_compliance_questions.htm.
15. 17 C.F.R. 275.204A-1—Investment Adviser Code of Ethics.
16. SEC, "SEC Names New Specialized Unit Chiefs and Head of New Office of Market Intelligence," January 13, 2010. As part of this program, the SEC also announced new tools available for facilitating cooperation with informants and witnesses, including cooperation agreements, deferred prosecution agreements, and non-prosecution agreements. In this manner, the SEC has increased incentives for cooperation and made it much more costly for potential wrongdoers to engage in their activities. See Robert S. Khuzani, "Remarks at News Conference Announcing Enforcement Cooperation Initiative and New Senior Leaders," January 13, 2010, www.sec.gov/news/speech/2010/spch011310rsk.htm.
17. See, for example, *SEC v. Balboa,* No. 11-cv-8731 (S.D.N.Y. Filed December 1, 2011); EC Litigation Release No. 22588, "Court Orders New York–Based Hedge Fund Manager and Firm to Pay Nearly $5 Million in Disgorgement and Penalties," January 9, 2013, available at www.sec.gov/litigation/litreleases/2013/lr22588.htm.
18. L. Richards, "Frequently-Asked Questions about SEC Examinations," SIFMA Compliance and Legal Division January General Luncheon Meeting, New York, January 17, 2008.
19. Rule 13d-1(c) under the 1934 Act.
20. Form PF can be found at www.sec.gov/about/forms/formpf.pdf.
21. Mary Jo White, "Remarks at the Securities Enforcement Forum," October 9, 2013, www.sec.gov/News/Speech/Detail/Speech/1370539872100#.VIzrmKTF9ew.
22. SEC, "SEC Announces Charges against Corporate Insiders Violating Laws Requiring Prompt Reporting of Transactions and Holdings," September 10, 2014, www.sec.gov/News/PressRelease/Detail/PressRelease/1370542904678.
23. This includes missing terms in investment agreements, such as deal fees, which the SEC is still considering banning.
24. A recent SEC case is instructive. In September 2014, the SEC charged a New York–based investment advisory firm with breaching its fiduciary duty to its private equity funds by commingling the funds in a manner that "improperly" benefited one over the other. For example, the two companies entered into joint lines of credit or had employees perform functions for both companies without sharing the cost between the two. As

a result, the investment firm agreed to settle with the SEC for more than $2.3 million. SEC, "SEC Charges New York–Based Private Equity Fund Adviser in Misallocation of Portfolio Company Expenses," September 22, 2014, www.sec.gov/News/PressRelease/Detail/PressRelease/1370543006673.

25. For example, a *Wall Street Journal* reporter looked at a review of regulatory filings from some 80 private equity firms and discovered that many made no mention of the consultants hired to manage a company or how partners were being compensated. Michael Wursthorn, "Private-Equity Consultants Face SEC Scrutiny: Regulators Seek Details on How Operating Partners Are Paid," October 8, 2014, www.wsj.com/articles/private-equity-consultants-face-sec-scrutiny-1412785084.
26. Andrew J. Bowden, "Spreading Sunshine in Private Equity," May 6, 2014, www.sec.gov/News/Speech/Detail/Speech/1370541735361.
27. Directive 2011/61/EU of the European Parliament and of the Council of 8 June 2011 on Alternative Investment Find Managers and Amending Directives 2003/41/EC and 2009/65/EC and Regulations (EC) No. 1060/2009 and (EU) No. 1095/2010 (hereafter referred to as "Official Journal"), (4), L174/1.
28. A fund that is domiciled in a single member state, operates and markets in its home state, and has no intention of managing an AIF or marketing outside its home state will still have to register and comply with the AIFMD. The EU passed the AIFMD into law in order to protect investors from a recurrence of the 2007–8 financial crisis, manage systemic risk, and create consistent rules for alternative investment managers. As a result of such regulatory consistency, the passport is a benefit to alternative fund managers. However, whether or not fund managers have any intention of using a passport, if they meet the requirements to register under the AIFMD, they will have to do so.
29. Official Journal (7), L174/2.
30. Official Journal (8), L174/2.
31. Official Journal, Chapter IV, Article 22(2), L174/32(a–f).
32. "Commission Delegated Regulation (EU) No. 231/2013 of 19 December 2012 supplementing Directive 2011/61/EU of the European Parliament and of the Council with regard to exemptions, general operating conditions, depositaries, leverage, transparency and supervision" is the formal name of the AIFMD Level II regulations (hereafter referred to as "Level II"), Section 4, Articles 47 and 48, L83/35.

    > *Registration and Light Regime Advice: AIFMD Level II specifies how the assets under management are to be calculated, how often, and the requirements for valuation. These funds must also make sure that they do not violate their light regime thresholds; therefore, if their assets exceed the amounts set for the light regime exemption, they must notify their local authority. If the exceeded amount is temporary, they may continue with their status; however, if the exceeded amount is constant for three months, the AIFM must apply for full AIFMD status within ninety days of the three-month period.*

    This comes from Level II, (5)(6)(7), L83/2.
33. Official Journal, Article 7(1)(2a–e), L174/20.
34. Level II, No. 231/2013 Section 4, Article 47 (1a–d), L83/34, 35.
35. Level II, No. 231/2013 Section 4, Article 47(2), L83/34, 35.
36. EMSA's technical advice to the European Commission on possible implementing measures of the Alternative Investment Fund Managers Directive, 2011/ESMA2011/379, VII, 212–13.
37. The AIFMD was created to protect investors' funds. In Article 49 of Level II is a chapter that is termed: "Alignment of investment strategy, liquidity profile, and redemption

policy." Level II goes into great detail regarding risk management. The risk management and accountability items addressed include the following:

Calculations and ongoing monitoring of assets under management

Calculations and requirements regarding liquidity

Calculation procedures of leverage

Professional liability and insurance requirements

Acting in the best interests of investors and the market

Due diligence

Conflicts of interest

Risk management systems

Separation of risk management functions

Accounting procedures

38. Financial Services Commission, "Information Page, Alternative Investment Fund Managers Directive Operating Conditions—Risk Management," 2.2.
39. Financial Services Commission, "Information Page, Alternative Investment Fund Managers Directive Operating Conditions—Risk Management," 7.1(a).
40. Ibid., 7.1(b).
41. Ibid., 7.1(c).
42. Official Journal, Chapter IV, Article 22, L174/32.
43. Level II, No. 231/2013, Article 110(3a–c), L83/63.
44. Official Journal, Chapter IV, Article 24.
45. Ibid., Article 26–30, L174/36–39.
46. Ibid., Article 30(1a–c), L174/38.
47. Veran Ross, "ESMA's Role in European and International Regulatory Cooperation," International Council of Securities Associations, Copenhagen, June 12, 2012, 10.
48. Official Journal, Chapter IX, Section 1, Article 45(1), L174/56.
49. Ibid., Article 45(1), L174/56.
50. Ibid., Article 46(1a–d), L174/57.
51. Ibid., Article 46(2a–m), L174/57–58.
52. Ibid., Article 45(4), L174/56.
53. Ibid., Article 45(5), L174/56.
54. Ibid., Article 45(6), L174/57.
55. Ibid., Article 45(11), L174/57.
56. Ibid., Article 45(7), L174/57.
57. Ibid., Article 45(8), L174/57.
58. Ibid., Article 48 (1), L174/59.
59. Ibid., Article 48 (2), L174/59.
60. Ibid., Article 49(1–2), L174/59.
61. Official Journal, Chapter IX, Section 1, Article 49(2), L174/59.
62. According to Patrick Rooney, regulatory affairs manager at the Irish Funds Industry Association, many non-EU funds continue to operate in the EU using "reverse solicitation." Reverse solicitation implies that non-EU managers do not actively market their funds or products in the EU, and they gain EU investors who learn about such funds and then solicit the non-EU manager. However, as Rooney states, reverse solicitation is not a sustainable model. It is supposed to be passive and not "marketing." However, the definition of *reverse solicitation* is left to the discretion of the member states, and different states have different interpretations. Some countries, such as France and Greece, are extremely

restrictive. If a fund is not AIFMD approved in those countries, it means there will be little or no tolerance of non-EU managers using reverse solicitation. France is considered to be a "closed shop." When the AIFMD was being drafted, France expressed its desire to restrict alternative asset managers to European managers only, and have no non-EU managers market their funds in the EU.

63. The SFC is one of four financial regulators in Hong Kong. The others are the Hong Kong Monetary Authority, which deals with banking; the Office of the Commissioner of Insurance; and the Mandatory Provident Fund Schemes Authority. The SFC is also responsible for enforcement of securities in Hong Kong and maintains a site dedicated to finding persons subject to arrest: www.sfc.hk/web/EN/regulatory-functions/enforcement/have-you-seen-these-people/people-subject-to-arrest-warrants.html.
64. While this section deals primarily with hedge funds, it should be noted that issues relating to transparency are relevant across a wide range of alternative investment products, particularly in the private equity market. By requiring disclosure of owners of corporations—including how many shares they may have—investors are better able to assess the type of product being offered. For more information, the SFC has prepared an outline concerning some of the required disclosures. It can be accessed here: http://en-rules.sfc.hk/net_file_store/new_rulebooks/h/k/HKSFC3527_4511_VER20.pdf.
65. Securities and Futures Commission, "Code on Unit Trusts and Mutual Funds," http://en-rules.sfc.hk/net_file_store/new_rulebooks/h/k/HKSFC3527_3039_VER20.pdf.
66. Generally, a fund doing business in Hong Kong must also nominate a local person to be approved by the SFC who can be served notifications of decisions by the commission.
67. MAS, "Overview," www.mas.gov.sg/About-MAS/Overview.aspx.
68. More detail, including the different requirements for funds of funds, can be found in Appendix 3 of Section II of the Code on Collective Investment Schemes, distributed by MAS.
69. www.fsa.go.jp/sesc/english/aboutsesc/actions.htm.
70. This ban began as a way to improve stability in the wake of continuing market uncertainty in 2008 and was not removed until 2013. See Takashi Nakamichi and Ayai Tomisawa, "Japan Cracks Down on Naked Short Selling," *Wall Street Journal*, October 28, 2008; Reuters, "Japan to Extend Naked Short Selling Ban to Oct," July 26, 2010.
71. Executive Bureau, Securities and Exchange Surveillance Commission, Inspection Manual for Financial Instruments Business Operators, II-1-5 Structures and Systems/Investment Management Business Operators, June 2014.

PART

6

Structured Products

CHAPTER 35

Structured Products I: Fixed-Income Derivatives and Asset-Backed Securities

The securitization of various financial assets provides liquidity, facilitates the transfer of risk, and presents unique risk-return opportunities to buyers of such securities. These products can also provide financing in ways that cannot be normally obtained through conventional loans. Structured products and credit risk derivatives were studied in Level I, where the following securities were analyzed: mortgage-backed securities (MBS), collateralized mortgage obligations (CMOs), credit derivatives and credit default swaps (CDSs), collateralized debt obligations (CDOs), and equity-linked products. Here, we expand the coverage of structured products by discussing fixed-income and interest rate derivatives, and asset-backed securities. A number of these alternative products were created in recent years and have been gaining increasing acceptance by market participants.

The first part of the chapter covers fixed-income derivatives, introducing the most important models of the term structure and then describing and valuing interest rates caps and floors, callable bonds, and interest rate swaps. The second part of the chapter describes and explains auto-loan-backed securities and credit card receivables.

35.1 OVERVIEW OF TERM STRUCTURE MODELING

Here, we briefly introduce the main approaches to modeling the term structure of interest rates.[1] These models belong to two groups: equilibrium and arbitrage-free. While theories have been proposed to explain the shape of yield curves (market expectations hypothesis, liquidity preference hypothesis, market segmentation hypothesis, and preferred habitat hypothesis),[2] the focus here is on introducing the models that have been developed to explain the evolution of bond prices. These models are then used to price fixed-income derivatives.

To price any interest-sensitive investment product, one needs to be able to model and explain the term structure of interest rates and its evolution through time. A simple example will highlight this point. Suppose we want to price a call option on a coupon bond that is free of credit risk. We already know that the most widely used model for pricing call options on common stocks is the Black-Scholes model, in which stock prices are assumed to have a lognormal distribution. Therefore, we might be tempted to apply the Black-Scholes model to bond prices by simply assuming that bond prices are lognormally distributed.

Let's, for now, ignore the issue of whether lognormal distributions are consistent with empirically observed distributions for bond prices. A more important problem that arises is how the distribution will change over time. After all, a 10-year bond becomes a nine-year bond after one year, an eight-year bond after another year, and eventually a one-year bond and then a one-day bond. While a 10-year bond is risky, with a price that fluctuates as interest rates change, the same bond becomes effectively riskless as it approaches maturity. This means that whatever we assume about the distribution of 10-year bond prices, we must also describe how that distribution will change over time. In other words, in order to be consistent, we must come up with return distributions for bonds belonging to all maturities.

In addition, we know that bond prices that are close to each other in terms of maturity are highly correlated, whereas bonds that are far apart are not highly correlated. In other words, we need to specify the correlations among bonds of different maturities as well. This gives the impression that modeling bond prices and their distributions is an impossible task. This would, in fact, be true if we didn't make some simplifying assumptions at the start. Therefore, it is important to bear in mind that the models and frameworks described next are highly simplified versions of the real world, and any attempt to make them more realistic will lead to very complex and perhaps unreliable models.

35.2 EQUILIBRIUM MODELS OF THE TERM STRUCTURE

Equilibrium models of the term structure (also referred to as first-generation models) make certain assumptions about the structure of fixed-income markets and then use economic reasoning to model bond prices and the term structure of interest rates. The first set of models introduced in this area took the process followed by the short-term interest rate as given. This was followed by the assumption that the unbiased expectation hypothesis holds for bond prices, meaning that the expected short-term rates of return on all bonds that are free from credit risk are assumed to be the same. This assumption, along with the exogenously specified process, was then used to model the entire term structure of interest rates. Here, we will introduce two of these models: Vasicek (1977) and Cox, Ingersoll, and Ross (1985).[3]

35.2.1 Vasicek's Model and Short-Term Interest Rates

Vasicek's model is a single-factor model of the term structure that assumes that the short-term interest rate drifts toward a prespecified long-term mean level. The model specifies the following mean-reverting process for the short-term rate of interest:

$$\tilde{r}_{t+1} = r_t + \kappa(\mu - r_t) + \sigma\tilde{\varepsilon}_{t+1} \tag{35.1}$$

This model states that next period's short-term rate, $\tilde{r}_{t+1}$, is equal to the current short-term rate, r_t, plus two adjustments. In the first adjustment term, κ and μ are constants and positive. According to the first adjustment term, $\kappa(\mu - r_t)$, the next period rate will be higher than the current rate if $\mu > r_t$. Therefore, one can think of

μ as the long-term average value of the short-term rate. This means that the short-term rate is likely to increase if it is currently below its long-term value and decrease if it is above the long-term value. This process is said to be mean-reverting, as the short-term rate tends to revert to its long-term mean. The speed of adjustment to the long-term rate is determined by the parameter κ. The higher its value κ, the faster the short-term rate will approach its long-term mean. The second adjustment factor, $\sigma\tilde{\varepsilon}_{t+1}$, introduces some noise into this adjustment process. The volatility of changes in interest rates is represented by σ and the noise is represented by $\tilde{\varepsilon}_{t+1}$, which is assumed to be a normally distributed random variable with a mean of zero and a standard deviation of 1. In terms of expected rates, Equation 35.1 can be derived as illustrated in Equation 35.2:

$$E[r_{t+1}] = r_t + \kappa(\mu - r_t) \tag{35.2}$$

In this model, the expected change in the short-term rate $E[r_{t+1} - r_t]$, is given by $\kappa(\mu - r_t)$. For example, if the current short-term rate is 5% and the long-term expected rate is 10%, the next period's short-term rate would be 5.5% for a κ value of 0.10 and 6% for a κ value of 0.2. Note that the unexpected change in the short-term rate is given by $\sigma \times \tilde{\varepsilon}$ and that the standard deviation of changes in the short-term rate is σ.

APPLICATION 35.2.1

Suppose the parameters of the Vasicek model are $\mu = 5\%$ (i.e., the long-term mean level of the short-term interest rate), and the following speed of adjustment parameter is given as $\kappa = 0.8$ and $\sigma = 1\%$. If the current short-term rate is 4%, what would be the expected short-term rate for the next period?

Inserting the given values into Equation 35.2 produces

$$E[r_{t+1}] = 0.04 + 0.8(0.05 - 0.04) = 4.8\%$$

If the current short-term rate is 6%, next year's short-term rate is expected to be 5.2%. Of course, Equation 35.2 has four variables, any one of which could be calculated given the other three.

In Vasicek's model, the volatility of changes in interest rates is constant and does not change as the level of interest rates change. In the previous example, the volatility of interest rate changes will be 1% regardless of the level of interest rates. This is one of the criticisms directed at Vasicek's model.

35.2.2 Vasicek's Model and the Term Structure of Interest Rates

In Vasicek's model, all bond prices are driven by one factor: the short-term interest rate. That is, the only source of uncertainty in the bond market is the random change in the short-term rate of interest. At this point, one may use several approaches to develop a mathematical formula for bond prices. For example, one could assume that the unbiased expectations hypothesis about the term structure holds. Under one form of this hypothesis, all bonds that are free from credit risk (regardless of their maturity) are expected to earn the same rate of return in the short run. Under this assumption, one can then obtain a mathematical formula for bond prices. Vasicek shows that the log of a bond price is given by:

$$\ln[B(t,T)] = A(t,T) - C(t,T) \times r_t \tag{35.3}$$

In Equation 35.3, the log of the current price of a bond that will mature at time T is explained by three terms. The first two terms, $A(t, T)$ and $C(t, T)$, are functions of time, whereas the last term is the current short-term rate. In particular, the value of $C(t, T)$ determines how bond prices will react to changes in the short-term rate and, therefore, how volatile the bond price is. The formula shows that $C(t, T)$ declines toward zero as the time to maturity declines. This is, of course, a desirable property of the model, since short-term bond prices are less sensitive to interest rate changes and display lower volatility. Equation 35.3 can also be used to represent the yield curve. Bond prices can be expressed as a function of the yield to maturity as $B(t, T) = \exp(-y(t, T) \times (T - t))$, where $y(t, T)$ is the yield to maturity. Therefore, the yield to maturity of the bond is given by

$$y(t,T) = -\frac{1}{T-t}[A(t,T) - C(t,) \times r_t] \tag{35.4}$$

This expression represents the yield curve, or the term structure of interest rates, in Vasicek's model.

35.2.3 Robustness of Vasicek's Model of the Term Structure of Interest Rates

The values of the three parameters in Vasicek's model can generate a downward-sloping, an upward-sloping, or a humped term structure. The term structure generated by Vasicek's model can be illustrated using the following example. Suppose that the current short-term interest rate is 5%, the long-term mean (μ) is 7%, the standard deviation is 1%, and the speed of adjustment (κ) is 0.1. Using Equation 35.4, this set of values for the parameters generates an upward-sloping term structure of interest rates, as illustrated in Exhibit 35.1 and Exhibit 35.2.

Exhibit 35.3 and Exhibit 35.4 illustrate the case of a downward-sloping yield curve generated by Vasicek's model. Exhibit 35.5 and Exhibit 35.6 illustrate the case of a humped yield curve.

35.2.4 The Cox, Ingersoll, and Ross Model of Interest Rates

Vasicek's model has been criticized because it allows the short-term interest rate to be negative. The **Cox, Ingersoll, and Ross model** (CIR model) alters the Vasicek model to make the variance of the short-term interest rate proportional to the rate itself,

EXHIBIT 35.1 Table of Vasicek's Model

Vasicek's Model		k	0.10	
		μ	0.07	
		σ	0.01	
		r	0.05	
$T - t$	$C(t,T)$	$A(t,T)$	$B(t,T)$	$y(t,T)$
0.5	0.4877	0.999	0.975	5.05%
1.0	0.9516	0.997	0.950	5.10%
1.5	1.3929	0.993	0.926	5.14%
2.0	1.8127	0.987	0.902	5.18%
2.5	2.2120	0.980	0.878	5.22%
3.0	2.5918	0.972	0.854	5.26%
3.5	2.9531	0.963	0.831	5.30%
4.0	3.2968	0.953	0.808	5.33%
4.5	3.6237	0.942	0.786	5.37%
5.0	3.9347	0.929	0.763	5.40%
5.5	4.2305	0.917	0.742	5.43%
6.0	4.5119	0.903	0.721	5.46%

thereby disallowing negative interest rates. In other words, negative rates are prevented because as rates approach zero, their volatility approaches zero. The following equation represents the short-term interest rate process proposed by the CIR model:

$$r_{t+1} = r_t + \kappa(\mu - r_t) + \sqrt{r_t}\sigma\tilde{\varepsilon}_{t+1} \tag{35.5}$$

where the three (constant) parameters κ, μ, and σ have the same meanings as before. The major difference between this model and Vasicek's is that the variance of the change in the short-term rate, $\sigma^2 r_t$, is proportional to the short-term rate. This makes sense, as we observe higher volatility for interest rate changes when the short-term rate is relatively high. The CIR model can generate yield curves with different shapes.

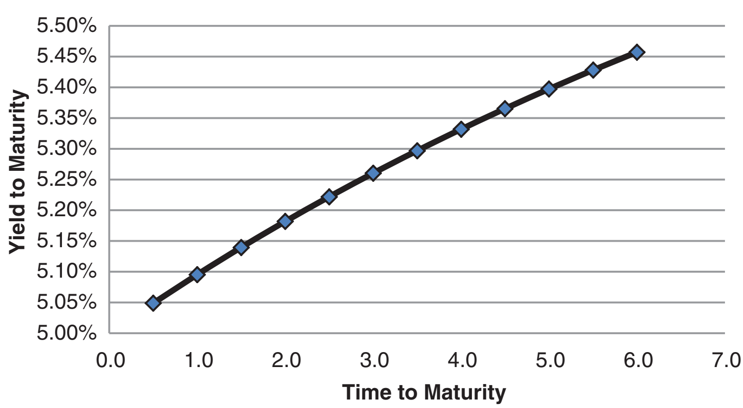


EXHIBIT 35.2 Vasicek's Model: Example of an Upward-Sloping Yield Curve

EXHIBIT 35.3 Vasicek's Model: Example of a Downward-Sloping Yield Curve

Vasicek's Model		k	0.10	
		μ	0.04	
		σ	0.01	
		r	0.06	
$T - t$	$C(t,T)$	$A(t,T)$	$B(t,T)$	$y(t,T)$
0.5	0.4877	1.000	0.971	5.95%
1.0	0.9516	0.998	0.943	5.90%
1.5	1.3929	0.996	0.916	5.85%
2.0	1.8127	0.993	0.890	5.81%
2.5	2.2120	0.989	0.866	5.76%
3.0	2.5918	0.984	0.842	5.72%
3.5	2.9531	0.979	0.820	5.67%
4.0	3.2968	0.973	0.798	5.63%
4.5	3.6237	0.967	0.778	5.59%
5.0	3.9347	0.960	0.758	5.54%
5.5	4.2305	0.952	0.739	5.50%
6.0	4.5119	0.944	0.720	5.46%

35.3 ARBITRAGE-FREE MODELS OF THE TERM STRUCTURE

Arbitrage-free models of the term structure (also referred to as second-generation models) use a different approach to model bond prices and the yield curve. For example, even if one starts with an exogenously given process for the short-term rate, a mathematical formula describing bond prices is obtained by arguing that the resulting bond prices should not allow for arbitrage opportunities. In particular, we know that under risk-neutral probability, the rates of returns on all investments—including bonds—should be equal to the short-term rate. Another distinguishing feature of

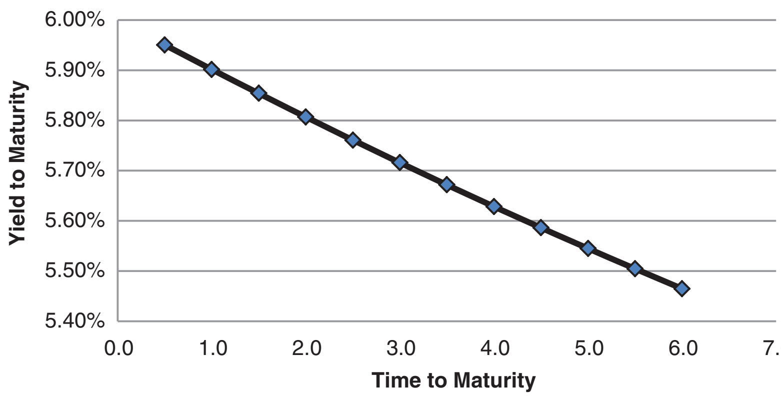


EXHIBIT 35.4 Vasicek's Model: Example of a Downward-Sloping Yield Curve

EXHIBIT 35.5 Vasicek's Model: Example of a Humped Yield Curve

Vasicek's Model		k	0.10	
		μ	0.06	
		σ	0.03	
		r	0.04	
$T - t$	$C(t,T)$	$A(t,T)$	$B(t,T)$	$y(t,T)$
0.5	0.4877	0.999	0.980	4.05%
1.0	0.9516	0.997	0.960	4.08%
1.5	1.3929	0.994	0.940	4.11%
2.0	1.8127	0.990	0.921	4.14%
2.5	2.2120	0.985	0.901	4.15%
3.0	2.5918	0.979	0.883	4.16%
3.5	2.9531	0.973	0.864	4.17%
4.0	3.2968	0.966	0.846	4.17%
4.5	3.6237	0.958	0.829	4.17%
5.0	3.9347	0.950	0.812	4.16%
5.5	4.2305	0.942	0.796	4.16%
6.0	4.5119	0.934	0.780	4.14%

arbitrage-free models is that the currently observed term structure is used to determine the parameters of the model. As a result, the theoretically derived term-structure model is consistent with the observed term structure. Therefore, any fixed-income derivative security that is priced using this theoretical model will be consistent with the current term structure and will also preclude arbitrage opportunities involving those derivatives and available bonds.

35.3.1 The Ho and Lee Model of Interest Rates

Ho and Lee (1986) proposed the first arbitrage-free model of interest rates. The **Ho and Lee model** assumes that the short-term interest rate follows a normally distributed process, with a drift parameter that is chosen so that the modeled interest

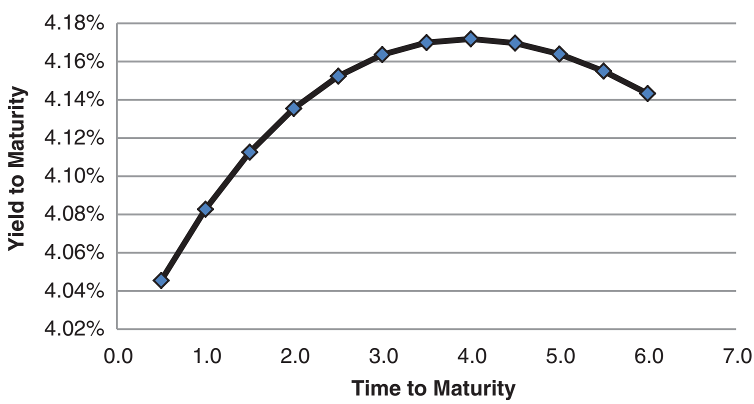


EXHIBIT 35.6 Vasicek's Model: Example of a Humped Yield Curve

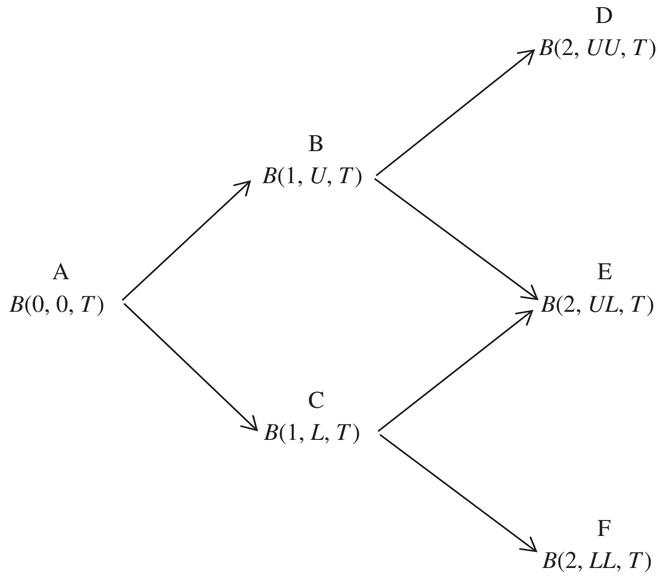


EXHIBIT 35.7 Two-Period Binomial Model for Bond Prices

rates fit the observed term structure of interest rates. Ho and Lee's model of the short-term rate is shown in Equation 35.6.

$$r_{t+1} = r_t + \theta_t + \sigma\tilde{\varepsilon}_{t+1} \tag{35.6}$$

where θ_t is a time-dependent mean change in the short-term rate (chosen to ensure that the model fits the initial term structure of interest), σ is the constant standard deviation of changes in the short rate, and $\tilde{\varepsilon}_{t+1}$ is a binomial random variable assuming values of +1 or −1. Notice that, unlike Vasicek's model, one of the parameters of Ho and Lee's model, θ_t, is not constant and is determined by the current term structure of interest rates.

35.3.2 The Ho and Lee Model in a Binomial Framework

Ho and Lee used this simple model as the basis of a binomial model for bond prices. The current prices are taken as given and are set up to match the currently observed term structure of interest rates. Thereafter, the term structure is assumed to be affected by random changes in interest rates. This is accomplished by allowing the whole term structure to evolve through time, as shown in the binomial model depicted in Exhibit 35.7.

Here, $B(t, i, T)$ is the price of a bond at time t and state i (which can be either U for "up" or L for "low") and with a face value of \$1 at time T. $B(0, 0, T)$ is today's observed term structure of bond prices. Bond prices evolve in response to random changes in interest rates. Ho and Lee use the well-known result that under risk-neutral probabilities the bond price in every state is equal to the expected value of the same bond in the next period discounted at the riskless rate. Using this result and the assumption about the source of uncertainty in the market, Ho and Lee are able to obtain analytical solutions for all bond prices in each future state.

35.3.3 Evaluation of the Ho and Lee Model of Interest Rates

Since the model is calibrated to fit the currently observed term structure of interest rates, the resulting prices for callable bonds, bond options, swaptions, and other interest rate derivatives will be consistent with the current term structure. In other words, the actions of arbitrageurs will be able to ensure that derivative prices are related to bond prices, so that arbitrage profits are not possible—hence the title of section 35.3: “Arbitrage-Free Models of the Term Structure.” Binomial interest rate trees will be used to value callable bonds in the next section.

The main disadvantages of the Ho and Lee model are that interest rates can be negative and that it assumes a very simple binomial process for bond prices. In the years since this model was devised, more sophisticated models that prevent negative interest rates while allowing for richer processes for bond prices have been developed.

35.4 INTEREST RATE DERIVATIVES

In this section, we describe and value the following interest rate derivatives: interest rate caps and floors, callable bonds, and interest rate swaps. The size of the interest rate derivatives market (gross notional outstanding figures) was \$504.4 trillion in December 2014 (Bank for International Settlements),[4] making it the largest derivatives market in the world. Also, according to the International Swaps and Derivatives Association (ISDA), 83% of the world’s top 500 companies used interest rate derivatives as of April 2009.

35.4.1 Interest Rate Caps and Floors

In an **interest rate cap**, one party agrees to pay the other when a specified reference rate is above a predetermined rate (known as the cap rate, which is similar to the strike price of a European call option). A **caplet** is an interest rate cap guaranteed for only one specific date. A **cap** is a series of caplets, and its price is equal to the sum of the prices of the caplets, which, in turn, can be valued using various term-structure models and a procedure similar to the Black-Scholes option pricing model.

Issuers of floating-rate debt can buy these options contracts to hedge against the possibility of increases in short-term interest rates (i.e., against variable or floating interest rate risk). Caps, also known as ceilings, work as insurance, a service for which purchasers of these contracts pay sellers a premium. Equation 35.7 denotes the periodic payment for a cap based on m periods per year:

$$\text{Cap Payment} = \text{Max}[(\text{Reference Rate} - \text{Strike Rate}), 0] \times \text{Notional Value}/m \tag{35.7}$$

To illustrate, consider a three-year interest rate cap. Party A buys the interest rate cap from party B with the following terms: The contract is for three years, the strike rate is 3%, the reference rate is three-month LIBOR, settlement is every three months, and the notional value is $10 million. Thus, every three months for the next three years, B will pay A if three-month LIBOR exceeds the strike rate of 3% at settlement. For example, suppose that the three-month LIBOR (i.e., the reference rate) is 4% on a settlement date. In this case, B will pay A $25,000, which is given by (4% – 3%) × $10,000,000/4 = $25,000. Suppose instead that three-month LIBOR is 2% on a settlement date. In this case, B will not make any payment to A. The maximum amount that A can lose from entering into this options contract is the up-front premium that A paid for the option.

APPLICATION 35.4.1A

Firm XYZ buys an interest rate cap from Bank DEF. The cap is for five years, has a strike rate of 5%, is settled quarterly, and has a notional value of $50 million. What are the payments, if any, from Bank DEF to Firm XYZ in the first four quarters if the reference rates for those quarters are, respectively, 4%, 5%, 6%, and 7%?

The solution is found using Equation 35.7, with $m = 4$ and the strike rate equal to 5%. For the third quarter, the formula is (6% – 5%) × $50,000,000/4, which is equal to $125,000. The four answers are $0, $0, $125,000, and $250,000. Note that the formula for a cap generates no payment when the strike rate equals or exceeds the reference rate.

In an **interest rate floor**, one party agrees to pay the other when a specified reference rate is below a predetermined rate (known as the floor rate, which is analogous to the strike price of a European put option). A **floorlet** is an interest rate floor guaranteed for only one specific date. A **floor** is a series of floorlets, and its price is equal to the sum of the prices of the floorlets. Similar to caps, floors can be valued using derivative pricing models like the Black-Scholes option pricing model.

These options contracts can be purchased by lenders in floating-rate debt to hedge against the possibility of declining short-term interest rates. A seller in an interest rate floor is compensated for guaranteeing the interest rate. Equation 35.8 denotes the periodic payment for a cap based on m periods per year:

$$\text{Floor Payment} = \text{Max}[(\text{Strike Rate} - \text{Reference Rate}), 0] \times \text{Notional Value}/m \tag{35.8}$$

To illustrate, consider a four-year interest rate floor. Party A buys the interest rate floor from party B with the following terms: The contract is for four years, the strike rate is 3%, the reference rate is three-month LIBOR, settlement is every three months, and the notional value is $20 million. Thus, every three months for the next four years B will pay A if three-month LIBOR is less than the strike rate of 3% at settlement. For example, suppose that the three-month LIBOR (i.e., the reference rate) is 1% on a settlement date. In this case, B will pay A $100,000, which is given by (3% – 1%) × $20,000,000/4 = $100,000. Suppose instead that three-month LIBOR is 4% on a settlement date. In this case, B will not make any payment to A. The maximum amount that A can lose from entering into this options contract is the up-front premium that A paid for the option.

APPLICATION 35.4.1B

Firm XYZ buys an interest rate floor from Bank DEF. The floor is for three years, has a strike rate of 7%, is settled quarterly, and has a notional value of \$10 million. What are the payments, if any, from Bank DEF to Firm XYZ in the first four quarters if the reference rates for those quarters are, respectively, 4%, 6%, 8%, and 10%?

The solution is found using Equation 35.8, with $m = 4$ and the strike rate equal to 7%. For the first quarter, the formula is (7% – 4%) × \$10,000,000/4, which is equal to \$75,000. The four answers are \$75,000, \$25,000, \$0, and \$0. Note that the formula for a floor generates no payment when the strike rate is equal to or less than the reference rate.

The parties to caps or floors are exposed to interest rate risk, which can be used to hedge the risk present in an existing position. However, since caps and floors are not typically exchange-traded instruments, there will be some exposure to counterparty risk. With either instrument, the buyers are exposed to counterparty risk. Because the sellers are paid up front, they have no exposure to the credit risk of the buyer.

35.4.2 Callable Bonds

Callable bonds give the bond issuer the right to redeem the bond by returning a prespecified payment to the investor during the period designated in the terms of the bond. Usually, a callable bond will become callable at a premium. For example, a bond that matures in 2035 might become callable on or after 2025 at a price of \$102 per \$100 of the principal amount. In the United States, most municipal bonds contain call provisions, almost all Treasury bonds and notes are non-callable, and many corporate bonds are callable. Callable bonds can also be found in Europe, Japan, Canada, Australia, Asia-Pacific, and Latin America.

Since the call feature is a call option held by the issuer, callable bonds can be valued using binomial interest rate trees, which depict a set of possible interest rate paths. For example, Exhibit 35.8 shows a two-period binomial interest rate tree.

In the tree, the interest rate for the first period is known today and is represented as $i(0, 0\text{–}1)$ in node A. This is the rate that will prevail between today and the end of the first period. If at the end of the first period we move upward to node B, this implies that we would have moved to the high interest rate state for the interest rate between the first and the second period, or $i(1, U, 1\text{–}2)$, where once again U stands for "up." If, on the other hand, at the end of the first period we move downward to node C, this implies that we would have moved to the low interest rate state for the interest rate between the first and the second period, or $i(1, L, 1\text{–}2)$, where L stands for "low." Changes in interest rates (for example, moving to node B or to node C)

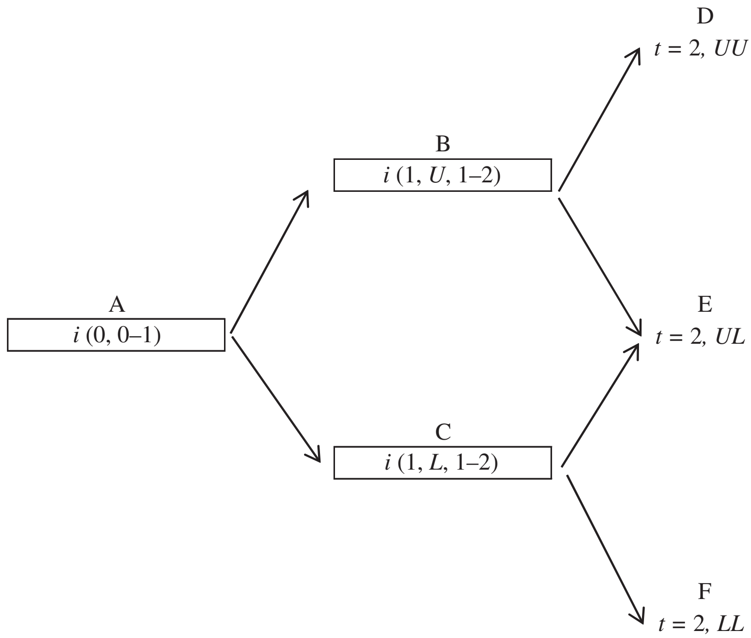


EXHIBIT 35.8 Two-Period Binomial Interest Rate Tree

are assumed to be random events. The bond matures at the end of the second period in nodes D, E, and F, and is valued using backward induction.

The construction of an interest rate tree is a complex task. The resulting tree should provide bond values that are arbitrage-free, as we explained in the section on yield-curve modeling. This can be accomplished when the set of interest rates belonging to each individual nodal period is related following an interest rate volatility assumption. For example, $i(1, U, 1\text{–}2)$ can be assumed to be related to $i(1, L, 1\text{–}2)$ as:

$$i\,(1, U, 1-2) = i\,(1, L, 1-2) \times e^{2\sigma} \tag{35.9}$$

where σ is the assumed volatility of the one-period interest rate.

APPLICATION 35.4.2A

Suppose a two-year bond pays an annual coupon of €5 and has a face value of €100. The one-year interest rate today is 3.56%, and one-year interest rates between the first year and the second year will be either 6.03% or 4.04%. The volatility of the one-year interest rate is assumed to be 20%. Notice that the interest rates between the first and second year are related, as depicted in Equation 35.9, because $i(1, U, 1\text{–}2) = 4.04 \times e^{2(0.20)} = 6.03\%$. Calculate the current value of the bond.

First, draw the binomial interest rate tree using the information provided in Exhibit 35.9.

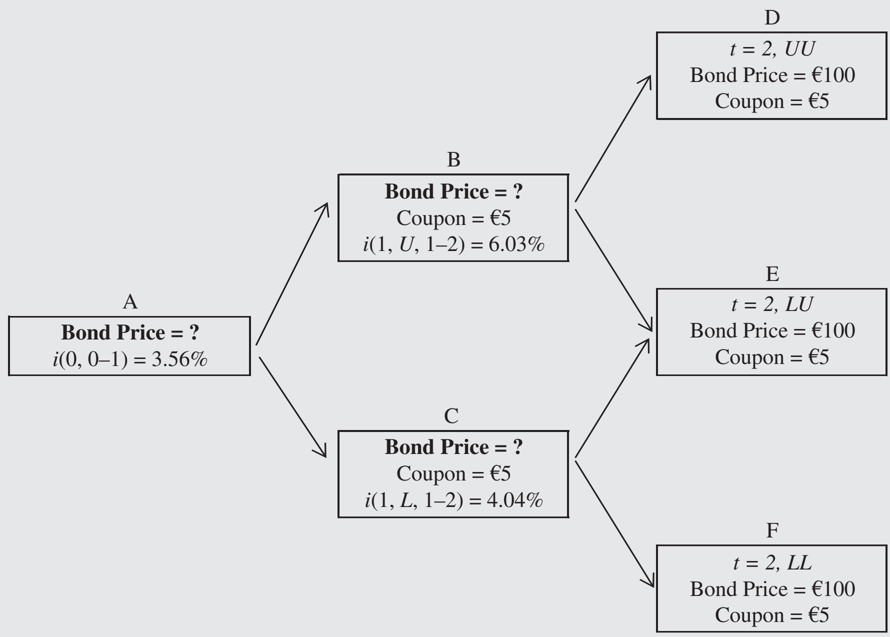


EXHIBIT 35.9 Two-Period Binomial Interest Rate Tree (Example 1)

Then, proceed to calculate bond prices at nodes B and C as follows (notice that these formulas assume that the probabilities of an up move and a down move are each 50%):

$$\text{Bond Price (Node B)} = \frac{1}{1 + 0.0603}[0.5 \times (100 + 5) + 0.5 \times (100 + 5)] = 99.029$$

$$\text{Bond Price (Node C)} = \frac{1}{1 + 0.0404}[0.5 \times (100 + 5) + 0.5 \times (100 + 5)] = 100.923$$

Finally, calculate the current value of the bond at node A:

$$\text{Bond Price (Node A)} = \frac{1}{1 + 0.0356}[0.5 \times (99.029 + 5) + 0.5 \times (100.923 + 5)]$$
$$= 101.367$$

Callability or other option features can be built into the valuation process of a binomial tree by recognizing the ability of an option owner to exercise the option at one or more nodes. For example, the option of a lender to call a bond issue for \$102 can be included in a binomial option analysis by replacing any valuation over \$102 with \$102 if the bond is callable at that point in time and if it is assumed that the lender will call the bond if its value is in excess of \$102.

APPLICATION 35.4.2B

Consider again a two-year bond that pays an annual coupon of €5 and has a face value of €100. The one-year interest rate today is 3.56%, and one-year interest rates between the first year and the second year will be either 6.03% or 4.04%. The volatility of the one-year interest rate is assumed to be 20%. Notice that the rates between the first and second year are related, as depicted in Equation 35.9, because $i(1, U, 1\text{–}2) = 4.04 \times e^{2(0.20)} = 6.03\%$. Now assume that this bond can be called at €100 in one year. What should the current price be?

In this case, the bond would be called in node C in year 1 because the market price of the bond at that node (€100.923) exceeds €100. The respective binomial interest rate tree, assuming the bond is called at \$100 in node C, has been drawn in Exhibit 35.10.

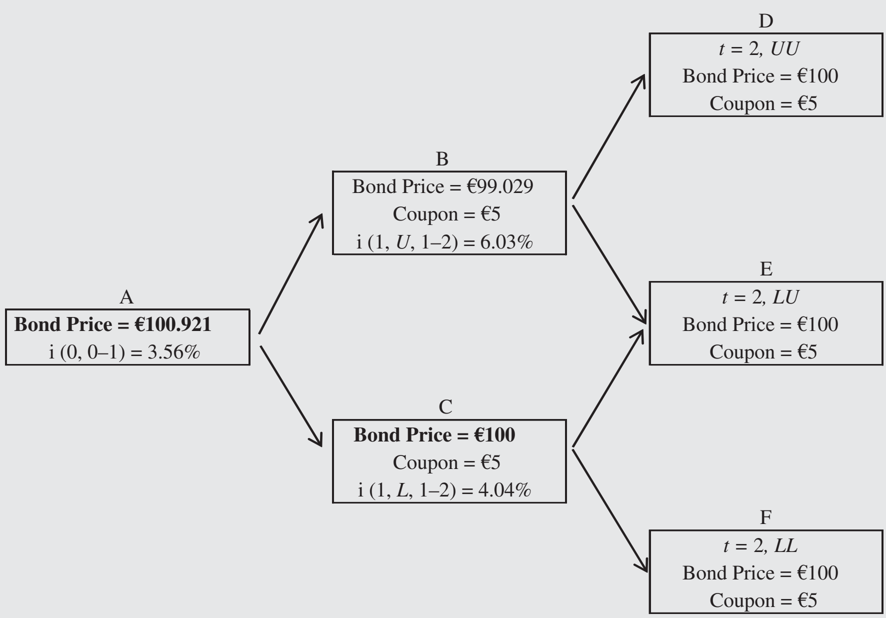


EXHIBIT 35.10 Two-Period Binomial Interest Rate Tree (Example 2)

$$\text{Bond Price (Node A)} = \frac{1}{1+0.0356}[0.5 \times (99.029 + 5) + 0.5 \times (100 + 5)]$$
$$= 100.921$$

Notice that, as expected, the current value of the callable bond (€100.921) is now lower than the current price of the non-callable bond (€101.367). The reason is that the bond would be called if in one year we move to the lower node C.

The difference between the two bond values in the examples is the value of the implicit or embedded call option that the issuer owns to call the bond:

$$\text{Value of Call Option} = \text{Value of Non-Callable Bond} - \text{Value of Callable Bond}$$

$$\text{Value of Call Option} = €101.367 - €100.921 = €0.446$$

In the preceding analysis, only the interest rate risk of the bond has been considered. Corporate and municipal callable bonds are exposed to credit risk as well. Therefore, the model should account for both interest rate and credit risk. Credit risk of a callable bond may play a role in the issuer's decision to call a bond. Calling a set of bonds will have an impact on the default probability of the remaining liabilities of the firm, as its cash balances as well as its liabilities will be reduced. Therefore, the firm may find it optimal not to call a bond that could be selling above its call price. In addition, a callable bond might also be convertible. In this case, the firm could decide to call the bond in order to force conversion. If investors decide to convert rather than redeem the bonds, the firm's equity will increase and the default risk of the remaining liabilities will be reduced.

35.4.3 Interest Rate Swaps

A swap is a contract between two parties to exchange cash flows at specified dates in the future according to prearranged rules. Swaps have already been described in the CAIA curriculum (for example, total return swaps were explained in the real estate section). In a "plain vanilla" **interest rate swap**, party A agrees to pay party B cash flows based on a fixed interest rate in exchange for receiving from B cash flows in accordance with a specified floating interest rate. Both payments are based on a notional principal and a specified number of years, which typically range from two to 15 years.

The payer in a vanilla swap is the party that agrees to pay a fixed rate in exchange for receiving a floating rate. The receiver (i.e., the counterparty) is the party that agrees to pay a floating rate in exchange for receiving a fixed rate. Interest rate swaps are subject to interest rate risk and credit risk (counterparty risk). Ignoring the counterparty risk, the payer in a vanilla swap will benefit from a rise in interest rates and will be hurt by a decline in interest rates. As a result, these instruments can be used to speculate, hedge, and manage interest rate risk. However, the most common motivation offered to explain the rationale of an interest rate swap is the comparative advantage argument.

For example, a firm may have a comparative advantage borrowing in the floating rate market but desires to borrow at the fixed rate. The firm then issues debt at the floating rate and, by entering into a swap contract with another party, is able to convert the floating-rate loan into a fixed-rate loan. Given the borrower's comparative advantage, the net fixed interest rate paid is lower after the swap transaction than would have been available by borrowing directly at the fixed rate. An interest rate swap can also be used to convert a liability from a fixed to a floating rate. It can also be used to convert an investment from a fixed to a floating rate, or from a floating to a fixed rate.

In recent years, pension funds have become active users of interest rate swaps. Pension funds' assets are managed to support liabilities that represent promises made to future retirees. In many countries, pension funds are expected to calculate the present value of these liabilities in order to determine the funding status of the fund. If the present value of liabilities exceeds the value of the pension fund's assets, then the fund may be considered to be underfunded. As a result, everything else being the same, a decline in interest rates will increase the present value of a fund's liabilities, increasing the gap between its assets and its liabilities. Pension funds have two broad options in managing this risk.

First, they could reduce the interest rate risk by investing in long-term bonds. In this case, a decline in interest rates will increase the values of their assets and liabilities, reducing the volatility in the gap between assets and liabilities. This strategy requires a pension fund to commit capital to the strategy, and therefore allocations to other asset classes have to be reduced.

Second, a pension fund may decide to invest its funds in asset classes that are expected to generate higher returns (e.g., private equity, hedge funds, or public equities) and then use an interest rate swap to manage its interest rate risk. In this case, the pension fund would agree to receive fixed payments in exchange for making floating payments. Should interest rates decline, the pension fund would benefit from a decline in the values of the future floating payments that it is expected to make.

Exhibit 35.11 illustrates the mechanics of an interest rate swap. Suppose that pension fund A has entered into an agreement to pay six-month LIBOR in exchange for receiving (from bank B) a fixed interest rate of 4% per annum every six months for four years, on a notional principal of $100 million. The table shows the resulting cash flows from the point of view of the pension fund, assuming the six-month LIBOR rates (expressed as rates per year with semiannual compounding) depicted in the second column of the table.[5]

On April 3 of year 1, the six-month LIBOR rate is 3.20%. This is the rate that would be applied to the floating payment made six months later, on October 3. Therefore, the first floating cash flow paid by the pension fund is equal to $1,600,000. This payment is calculated as follows: (3.20%/2) × $100,000,000 = $1,600,000. The same procedure can be followed to find the floating rate payments that will be made in subsequent periods. The net cash flow to the pension fund is equal to the difference between the fixed cash flow to be received and the floating cash flow to be paid. The principal in a swap contract (known as notional principal) is not exchanged at

EXHIBIT 35.11 Interest Rate Swap Example

Date	Six-Month LIBOR	Floating Cash Flow	Fixed Cash Flow	Net Cash Flow
April 3, year 1	3.20%			
October 3, year 1	3.50%	–$1,600,000	$2,000,000	$400,000
April 3, year 2	4.00%	–$1,750,000	$2,000,000	$250,000
October 3, year 2	4.50%	–$2,000,000	$2,000,000	$ 0
April 3, year 3	4.60%	–$2,250,000	$2,000,000	–$250,000
October 3, year 3	4.10%	–$2,300,000	$2,000,000	–$300,000
April 3, year 4	3.90%	–$2,050,000	$2,000,000	–$ 50,000
October 3, year 4	3.70%	–$1,950,000	$2,000,000	$ 50,000
April 3, year 5		–$1,850,000	$2,000,000	$150,000

the end of the life of the swap; it is used only for the computation of interest payments. In practice, only the net cash flows, or the difference between the fixed and floating rate payments, are exchanged.

The fixed rate of an interest rate swap is referred to as the **swap rate**. Initially, the swap rate is set so that the present value of expected floating payments is equal to the present value of expected fixed payments. Ignoring counterparty risks, the expected fixed payments are known with certainty and the expected floating payments can be estimated from the currently available interest rate futures prices. Then, using all available information, the swap rate is set so that the present values of fixed and floating payments are equal. An example later in this section will demonstrate this procedure.

Similar to other fixed-income instruments, swaps of different maturities carry different swap rates. By the same token, the **swap rate curve** displays the relationship between swap rates and the maturities of their corresponding contracts, having a concept analogous to that of the yield curve. The swap rate curve is an important benchmark for interest rates in the United States. It is also frequently used in Europe as the benchmark for all European government bonds.

Worldwide, interest rate swaps represent one of the most important interest rate derivative contracts in terms of size and volume. More specifically, total volume in cleared interest rate swaps reached $388.5 trillion (notional amount) in 2014. Dealer-to-dealer clearing and dealer-to-client clearing were $266.1 trillion and $57.4 trillion, respectively. Euro-denominated swaps accounted for nearly 40% of volume, followed by dollar-denominated swaps (29.2% of volume). Open interest at the end of 2014 was $181 trillion (*Futures Industry* 2015).

35.4.4 Valuation of an Interest Rate Swap

Interest rate swaps are worth zero when the two parties agree to the transaction. Once the contract is entered into, payments from the floating-rate party or leg of the agreement will change as market interest rates change.

An interest rate swap is equivalent to a bond transaction in which the fixed-rate payer issues a fixed-coupon bond and invests the proceeds in a floating-rate bond with the same payment dates and maturity. Then, on each payment date, the floating-rate payment is received and the fixed-coupon payment is made. Thus, the swap can be valued as the difference between the market value of the fixed-coupon bond and the market value of the floating-rate bond. It is important to bear in mind that interest payments are netted in the actual swap, and that the contract does not require principal payments. The procedure of estimating the market values of fixed- and floating-rate bonds is simply an artifice that facilitates the calculation of the value of the swap.

The valuation of an interest rate swap will be explained using an example.[6]

APPLICATION 35.4.4A

On January 1, ABC pension fund enters a one-year swap, agreeing to pay 4.3464% fixed rate on a notional amount of $10 million and receive a floating payment based on three-month LIBOR. Both the fixed and the floating payments will be made on a quarterly basis. The three-month LIBOR rate on January 1 is observed to be 4%. In addition, the interest rate futures market indicates the following rates for the next three quarters: 4.20%, 4.40%, and 4.80%. Calculate the expected payments for the swap.

The expected payments for both fixed and floating payments are displayed in Exhibit 35.12.[7]

EXHIBIT 35.12 Fixed and Floating Payments

(1)	(2)	(3)	(4)	(5)	(6)	(7) = (6) × 10M	(8) = 4.3464% × (3)/360 × 10M
Quarter Starts	Quarter Ends	Number of Days in Quarter	Current LIBOR	Future LIBOR Rates Start of Quarter	Quarterly Future LIBOR Start of Quarter	Floating Payment End of Quarter	Fixed Payment End of Quarter
January 1	March 31	90	4.00%		1.00%	100,000	108,660
April 1	June 30	90		4.20%	1.05%	105,000	108,660
July 1	September 30	90		4.40%	1.10%	110,000	108,660
October 1	December 31	90		4.80%	1.20%	120,000	108,660

Notice that floating payments are made at the end of each quarter based on the three-month LIBOR rate observed at the beginning of the same quarter. For instance,

$$100{,}000 = \frac{90}{360} \times 4.0\% \times 10{,}000{,}000$$

$$105,000 = \frac{90}{360} \times 4.2\% \times 10{,}000{,}000$$

Similarly, the fixed payments are calculated using the swap rate of 4.3464%. For instance,

$$108{,}660 = \frac{90}{360} \times 4.3464\% \times 10{,}000{,}000$$

A question that arises from this example is why the swap rate is set equal to 4.3464%. As discussed, given all available information, this is the rate that sets the present value of future fixed payments equal to the present value of future floating payments. In other words, 4.3464% is the swap rate that sets the net value of the swap to zero. After the swap agreement has been made, market interest rates will change and the swap's value will vary. The next example demonstrates the valuation of the swap agreement used in the previous example.

APPLICATION 35.4.4B

Given the cash flows and interest rates from Application 35.4.4A, calculate the value of the swap as the discounted values of the expected cash flows. To value the expected future cash flows of the swap, it is necessary to specify the discount rate that needs to be applied to future cash flows. It turns out that the interest rates obtained from the futures contracts can provide us with the information needed to calculate these present values. Exhibit 35.13 is based on the figures displayed in Exhibit 35.12, but three new columns have been added and columns 3–5 have been removed for space concerns. The exhibit displays

EXHIBIT 35.13 Present Values of Fixed and Floating Payments

(1)	(2)	(6)	(7)	(8)	(9)	(10) = (7) × (9)	(11) = (8) × (9)
Quarter Starts	**Quarter Ends**	**Quarterly Future LIBOR Start of Quarter**	**Floating Payment End of Quarter**	**Fixed Payment End of Quarter**	**Forward Discount**	**PV of Floating Payments**	**PV of Fixed Payments**
January 1	March 31	1.00%	100,000	108,660	0.990099	99,010	107,584
April 1	June 30	1.05%	105,000	108,660	0.979811	102,880	106,466
July 1	September 30	1.10%	110,000	108,660	0.969150	106,607	105,307
October 1	December 31	1.20%	120,000	108,660	0.957658	114,919	104,059
Total						**423,416**	**423,416**

Note: There is a slight rounding error in the last column of this exhibit.

all the information needed to calculate the present values of the two streams of cash flows.

Note that the sum of the present values of the floating payments in column 10 is equal to the sum of the present values of the fixed payments in column 11. This demonstrates that the swap has an initial value of zero. Note that a twelfth column of the netted expected cash flows could be formed and used to calculate the same net value.

The keys to our calculations are the forward discounts that appear in column 9. They are based on the quarterly three-month LIBOR rates that appear in column 6. For instance,

$$0.990099 = \frac{1}{(1 + 1.00\%)}$$

$$0.979811 = \frac{1}{(1 + 1.00\%) \times (1 + 1.05\%)}$$

In other words, the denominator of each discount factor is compounded using the current and previous three-month LIBOR rates. The present values of the two streams are calculated using these discount rates. It can be seen that when using 4.3464% as the swap rate, the present values of the two streams are equal when the swap contract is initiated.

As an additional example, suppose that after the first quarterly payments interest rates increase, changing the current and remaining three-month LIBOR rates to 4.4%, 4.8%, and 5.0%, respectively. What is the gain by the fixed-rate payer?

Exhibit 35.14 and Exhibit 35.15 duplicate the calculations displayed in Exhibit 35.12 and Exhibit 35.13 using the new three-month LIBOR futures, taking into account that the first payments have already been made.

It can be seen that while the fixed payments remain the same, the floating payments have increased, benefiting the party that pays the fixed rate. Given the new structure of the three-month LIBOR rates, we can calculate the present values of the two streams.

EXHIBIT 35.14 Interest Rate Swap Payments after a Change in Three-Month LIBOR

(1) Quarter Starts	(2) Quarter Ends	(3) Number of Days in Quarter	(4) Current LIBOR	(5) Future LIBOR Rates Start of Quarter	(6) Quarterly Future LIBOR Start of Quarter	(7) Floating Payment End of Quarter	(8) Fixed Payment End of Quarter
April 1	June 30	90	4.40%		1.10%	110,000	108,660
July 1	September 30	90		4.80%	1.20%	120,000	108,660
October 1	December 31	90		5.00%	1.25%	125,000	108,660

EXHIBIT 35.15 Present Values of Fixed and Floating Payments

(1) Quarter Starts	(2) Quarter Ends	(6) Quarterly Future LIBOR Start of Quarter	(7) Floating Payment End of Quarter	(8) Fixed Payment End of Quarter	(9) Forward Discount	(10) PV of Floating Payments	(11) PV of Fixed Payments
April 1	June 30	1.10%	110,000	108,660	0.989120	108,803	107,478
July 1	September 30	1.20%	120,000	108,660	0.977391	117,287	106,203
October 1	December 31	1.25%	125,000	108,660	0.965324	120,666	104,892
Total						**346,756**	**318,573**

The present value of the remaining floating payments will be higher than the present value of the remaining fixed payments, benefiting the fixed-rate payer. In other words, while the net present value (NPV) of the swap was zero when it was initiated, the NPV became positive for the fixed-rate payer (i.e., \$28,183 = \$346,756 −\$318,573), and negative (i.e., −\$28,183) for the floating-rate payer.

35.4.5 Risks Associated with Interest Rate Swaps

In this section, we briefly discuss the main risks to which interest rate swaps are subject—namely, credit risk and interest rate risk. Furthermore, the events of 2007–8 showed that it is no longer acceptable to assume that top-tier banks could never default. As a consequence, LIBOR rates should not be regarded as risk-free rates, a problem that in turn affects the valuation of interest rate swaps.

35.4.5.1 Credit Risk Credit risk on a two-leg swap exists when one of the parties to the contract is in-the-money, because that leg of the contract will face the possibility of default by the other party. On the other hand, when a swap is agreed upon through an intermediary (i.e., a financial institution), typically the intermediary will bear the default risk in exchange for a fixed percentage of the value of the contract in the form of a bid-ask spread.

The credit risk of an interest rate swap can be managed according to two dimensions: (1) contractual provisions, documentation, collateral, and contingencies and (2) diversification of the swap book across industry and market sectors (Sundaresan 2002).

35.4.5.2 Interest Rate Risk The risk exposure of a swap due to unanticipated interest rate changes is another potentially important risk. For example, Ferrara and Ali (2013) simulate many forward yield curves (using an arbitrage-free interest rate model), and evaluate the potential exposure of vanilla interest rate swaps under the most familiar yield curve shapes and under different volatility assumptions. The authors highlight that unanticipated changing interest rates can, on one hand, create substantial mark-to-market (MTM) or counterparty exposure, which may cause significant MTM losses and require substantial collateral posting. On the other hand, they also find that unanticipated changing interest rates can generate considerable

MTM gains, which can lead to counterparty exposure if the swap contract is not collateralized.

Credit risk and interest rate risk interact in fine ways. These interactions can be examined by estimating the MTM value of swaps for a range of term-structure scenarios and credit-risk assumptions. These estimations can be performed using Monte Carlo simulation or other techniques.

35.4.5.3 The Situation after the Financial Crisis of 2007–8 The two key assumptions under which the traditional approach to pricing and valuing standard interest rate swaps is based are that LIBOR discount factors are (1) reasonable proxies for the credit quality of the counterparty when the contract is uncollateralized, and (2) suitable measures for the risk-free term structure when the contract is collateralized.

Smith (2012) argues that the financial crisis of 2007 defies the second assumption. This is because collateralization is now usual in the swap market, and the existence of considerable and persistent differences between LIBOR and other proxies for risk-free rates implies that LIBOR discount factors can no longer be regarded as risk-free rates. Because of this, fixed rates on overnight indexed swaps are now considered more appropriate for valuing collateralized contracts. The spread can arise in two ways—first, as a liquidity premium to compensate for liquidity risk, and second, as a credit spread.

Prior to the 2007–8 global financial crisis, most swap market participants ignored the counterparty risk associated with large global banks. The reason was that most assumed that these institutions would never default on their obligations. The global financial crisis changed all of that, and as a result, a credit spread reflecting the counterparty risk is now incorporated into swap spreads. In this sense, one should consider the events that took place during the global financial crisis as observing a black swan for the very first time. It forever changed the probability of observing black swans from zero to some meaningful figure (see Davis 2015).

This section on the risks associated with interest rate swaps ends by analyzing the case of the swaps entered by Harvard University in 2004. Interest rates at that time were historically low, and Harvard entered into a series of interest rate swaps in anticipation of the future funding needs for the construction of the Allston Science Center project (planned to begin in 2007), as well as other projects. Harvard had issued variable interest rate debt and desired to hedge against the risk of rising interest rates, as well as on expected future debt issuances. According to Ferrara and Ali (2013), who summarize the problems that the interest rate swaps caused Harvard (first in 2005 and then after the substantial decline in interest rates following the financial crisis of 2008), the notional values of interest rate swaps held by Harvard had increased to $3.7 billion by the end of 2005. Harvard would have been forced to pay $461.2 million to exit these positions. The notional values of interest rate swaps held by Harvard fluctuated between 2005 and 2008, and stood at $3.5 billion by the end of 2008. Since Harvard had entered these swaps as payer of the fixed leg, the subsequent significant drop in interest rates caused the values of these interest rate swaps to experience significant declines, negatively affecting Harvard's financial position.

Harvard's problems with its interest rate swaps were compounded by the fact that the swaps required the university to deliver collateral (in cash) proportional to the magnitude of the NPV on its interest rate swaps, which in this case had become negative. As Harvard's NPV became increasingly negative, the amount of cash it had to post as collateral increased, thus creating an illiquidity problem for the university. These difficulties explain, at least partially, Harvard's decision to pay almost $500 million in fiscal year 2009 to terminate a subset of its portfolio of interest rate swaps that had a total notional value of $1.1 billion.

35.5 ASSET-BACKED SECURITIES

Structured products were introduced in Level I, where it was argued that the concept of structuring is essential to finance and investments. Structuring is the practice of engineering unique financial opportunities from existing asset positions. This process enables diverse investors to hold claims with different risk exposures (or other characteristics) from the same assets. To create asset-backed securities (ABS), an entity creates a special purpose vehicle (SPV), to which it sells the respective assets. Thereafter, financial securities are issued from the SPV to investors. Most asset-backed securities are issued in three separate tranches (classes A, B, and C). The highest tranche (class A, or senior tranche) is typically the largest tranche, has the highest credit rating, and is supported by the junior tranches. The B and C tranches have lower credit ratings and thus earn higher yields to compensate investors. The lower tranches suffer most or all of the losses when a credit event occurs. However, if the losses are large enough, the A tranche will also experience losses.

In a sequential-pay tranche structure, the senior tranche receives all initial principal payments (both scheduled and prepaid) until all tranche holders have been completely paid off. After that, the next most senior tranche is entitled to receive all the principal payments, and so on. Each tranche will collect interest payments as long as the tranche's principal has not been paid off. Principal payments are distributed to each tranche according to a predetermined priority scheme. Each tranche has a different average maturity, to adapt to different investor clienteles.

This section introduces two types of asset-backed securities: auto loans and credit card receivables. From an investor's perspective, the shorter duration and amortizing plans of asset-backed securities facilitate a reduction in portfolio interest rate and possibly credit risk. Exhibit 35.16 shows the total amount of U.S. asset-backed securities outstanding by type.[8] We will comment on the statistics presented in this exhibit as we progress through the section.

In the case of Europe, the aggregate ABS market experienced phenomenal growth during the decade leading up to 2008, peaking before the global financial crisis of 2008–9, with a total of $1.2 trillion ABS issued in 2008 (this includes residential mortgage-backed securities, commercial mortgage-backed securities, securities backed by loans to small and midsize enterprises, CDOs, and the ABS discussed here). By 2013, total new issuance was only $239 billion. According to Altomonte and Bussoli (2014), demand for these assets plunged after the crisis because of the worsening of the quality of the collateral underlying the various categories of ABS. Furthermore, the post-crisis regulatory environment accentuated the market decline, as ABS products were blamed for creating and amplifying the crisis. In the post-crisis period, new and stricter capital requirements became mandatory for the issuance of new ABS.

EXHIBIT 35.16 U.S. Asset-Backed Securities Outstanding ($ Billions)

Year	Automobile	Credit Card	Housing-Related	Student Loans	Other	Total
1985	0.9	0.0	0.0	0.0	0.3	1.2
1986	10.5	0.0	0.0	0.0	0.7	11.2
1987	14.2	2.4	0.0	0.0	1.5	18.1
1988	13.5	9.1	0.0	0.0	3.1	25.7
1989	14.1	20.0	0.0	0.0	2.8	36.9
1990	19.3	42.1	0.0	0.0	3.6	65.0
1991	27.2	59.0	0.0	0.2	4.0	90.4
1992	36.8	70.8	0.0	0.4	7.0	115.0
1993	42.3	75.1	0.0	0.8	12.2	130.4
1994	39.3	98.2	0.0	3.4	19.1	160.0
1995	52.8	129.9	0.0	6.5	23.2	212.4
1996	66.7	167.1	0.0	14.3	40.2	288.3
1997	79.4	191.0	0.0	25.9	55.4	351.7
1998	88.5	199.6	0.0	31.5	70.5	390.1
1999	109.4	213.8	0.0	36.4	88.1	447.7
2000	140.5	236.8	0.0	44.7	103.1	525.1
2001	167.0	265.9	0.0	48.1	111.8	592.8
2002	187.6	293.3	0.1	58.7	106.6	646.3
2003	191.5	303.5	0.1	87.8	114.2	697.2
2004	177.3	297.5	0.3	122.5	115.6	713.2
2005	195.9	287.2	0.5	159.6	129.1	772.3
2006	196.2	291.5	0.7	200.6	147.9	836.8
2007	181.2	324.4	0.5	229.6	164.6	900.3
2008	140.4	315.6	0.2	237.9	149.8	844.0
2009	127.6	300.3	2.2	239.5	142.6	812.1
2010	115.9	216.8	3.6	240.6	130.0	706.9
2011	117.2	163.9	3.4	234.2	128.8	647.5
2012	142.4	127.9	3.8	233.5	139.4	647.0
2013	161.5	124.5	8.7	228.4	152.9	676.0
2014	178.9	136.5	25.6	216.5	166.8	724.2

Note: This table does not include CDOs.
Source: Securities Industry and Financial Markets Association (SIFMA) (2014) and author's calculations.

35.6 AUTO LOAN–BACKED SECURITIES

Auto loan–backed securities (ALBS) receive cash flows from customer payments assembled from a specific pool of auto loans or leases. They can be issued by commercial banks, financial subsidiaries of auto companies, independent finance companies, or small financial companies dedicated to auto loans. Auto loan (and auto lease) asset-backed securities have historically represented between a fifth and a quarter of the asset-backed securities market in the United States. Total auto loan–backed securities outstanding in the United States amounted to $178.9 billion in 2014, reaching a new post-crisis record (the all-time record occurred in 2006).

A borrower's interest rate is affected by income, employment history, credit scores (categorized as prime, nonprime, or subprime), and other variables. The cash flows of auto loan–backed securities are managed by servicers, which are paid a fixed fee (usually 0.5% to 2.0% of the pool's remaining collateral balance). Servicers collect and process borrowers' payments to the loan pool and then pay investors in this asset-backed security. They also endeavor to collect payments due from delinquent borrowers. The credit rating of auto loan–backed securities is affected by the creditworthiness of the servicer and also by its experience.

Auto loan–backed securities receive cash flows from monthly loan payments (interest and scheduled principal repayments) and from prepayments made by borrowers. In general, the rate of prepayments on asset-backed securities will be affected by a number of factors. The first is the current level of interest rates relative to the rates being charged on the loans in the pool. Borrowers will have a greater incentive to refinance when current loan rates are low relative to the rates on existing loans. Refinancing, however, is not an important factor explaining prepayments in the case of auto loans. This is both because auto loans are generally small and because the value of a vehicle may depreciate at a faster pace than the outstanding balance of the loan in the first few years. Prepayments nonetheless can also originate due to the following factors: sales and trade-ins involving full loan payment, loss or destruction of an automobile, repossession and subsequent resale of a vehicle, and payoff of the loan by the borrower with cash.

A prepaid loan provides an investor in this asset-backed security with early return of principal but takes away the investor's ability to earn the original interest rate on the old loan. Similarly, if a borrower does not fully prepay a loan but merely makes additional (unscheduled) partial principal repayments, investors in auto loan–backed securities experience effects similar to those of the full prepayment scenario but at a smaller magnitude. As a result, prepayments affect not only the cash flows relating to the principal but also the interest payments.

Prepayments create an element of uncertainty in the cash flows received by investors in auto loan–backed securities. Analysts build fundamental models of prepayment speeds and analyze past prepayment rates in their effort to predict future prepayment rates. It is thus important that investors take into account the speed of prepayments, which can be measured by the conditional prepayment rate (CPR) and calculated using the following formula, based on annualizing a monthly prepayment or single monthly mortality (SMM) rate (Fabozzi and Mann 2012):

$$\text{CPR} = 1 - (1 - \text{SMM})^{12} \tag{35.10}$$

Historically, SMMs have increased as loans season. The CPR method is used to measure prepayments for most asset-backed securities. However, in the case of auto loans, the convention is to use the **absolute prepayment speed (ABS)**, which is defined as the monthly prepayment expressed as a percentage of the original collateral amount. The following formula relates the SMM and ABS measures:

$$\text{SMM} = \frac{\text{ABS}}{1 - [\text{ABS} \times (M - 1)]} \tag{35.11}$$

where M is the number of months after the loan originated.

Example: If the ABS rate is 1.3% at month 12 after origination, calculate the SMM.

Answer:

$$\text{SMM} = \frac{0.013}{1 - [0.013 \times (12 - 1)]} = 0.0152 = 1.52\%$$

The ABS can also be expressed in terms of SMM:

$$\text{ABS} = \frac{\text{SMM}}{1 - [\text{SMM} \times (M - 1)]} \tag{35.12}$$

Auto loan–backed securities do not have government backing and are therefore subject to credit risk. Auto loans are typically classified as prime, near-prime, or subprime. The credit risk and performance of ALBS depend on the structure of the ABS and the performance of the underlying auto loan collateral. Auto loan–backed securities based on subprime loans usually include a provision that protects either investors in the senior ABS tranches or the entire structure. In the case of subprime auto loan–backed securities, one common credit enhancement is overcollateralization. Moreover, most subprime ALBS contain a credit enhancement provision that redirects any interest received on the underlying auto loan portfolio in excess of interest payments on junior ABS tranches to a reserve account, which offers an additional credit safeguard planned to benefit senior tranches (Culp and Forrester 2015). Finally, the quality of the originator's underwriting standards also influences an investor's decision to invest in a specific ALBS.

Prepayment risk is relatively low for auto loan–backed securities. Heitfield and Sabarwal (2004) find that prepayment rates for ALBS are generally not affected by prevailing market interest rates but increase rapidly with loan age. They also find that ALBS default rates are much more vulnerable to aggregate economic shocks than are prepayment rates. Furthermore, default rates are usually preceded by increases in unemployment rates.

Finally, liquidity and mark-to-market risks will generally be greater for ALBS, as with all credit markets, during periods of heightened risk aversion.

35.7 CREDIT CARD RECEIVABLES

When a customer makes a purchase on a credit card, the issuer (i.e., the lender) of the credit card provides credit to the cardholder (i.e., the borrower). The sum that the cardholder owes the lender is a receivable to the credit card issuer. Credit card debt does not have an actual maturity date, because principal repayment is not scheduled (i.e., it is regarded as a non-amortizing loan). A **credit card receivable (CCR)** is an asset-back security in which a pool of credit card receivables is used as collateral. The cash flows received by CCRs consist of annual fees, interest, and principal payments. The total amount of CCRs outstanding in the United States was \$136.5 billion in 2014, down more than 50% from an all-time high of \$324.4 billion in 2007.

CCRs are structured differently than other asset-backed securities (such as those based on auto loans and mortgages). This is because credit card receivables, which have a short-term life span and tend to be paid off within a year (i.e., much faster than an auto loan or a mortgage), support the outstanding certificates issued by the trust, which usually have maturities of three, five, or 10 years.

Furthermore, credit cards ought to meet monthly payments, but since loan balances are revolving, the principal is not amortized. This characteristic led to the development of a structure, the master trust, which is better adapted to the short-term nature of CCRs because it allows issuers to sell more than one series from the same pool of credit card receivables. This is accomplished by including more receivables every time a new series is sold. As a result, the loan balance may not need to be paid off completely.

Because of this, CCRs are structured so as to pay periodic interest to their holders, but no principal is paid during a prespecified lockout period (which may last anywhere from a year and a half to 10 years). If credit card users make principal payments during the lockout period, these cash flows are used to pay for new credit card receivables, with the aim of maintaining a relatively constant overall value of the receivables. Once the lockout period is over, principal payments are channeled to CCR holders in what is termed the principal amortization period.

The credit risk of the underlying collateral in the master trust can change over time because credit card issuers can add receivables to the trust and also change the terms on existing receivables. This requires rating agencies to constantly examine the performance of the receivables in the trust.

Credit cards and auto loans are **recourse loans**, which means that the borrower is personally liable for repaying any outstanding balance on the loan. This implies that lenders are allowed to garnish wages or levy accounts to collect what is owed, even after they have taken collateral. On the other hand, in a **non-recourse loan**, the lender can collect only the collateral at hand.

Credit enhancements can be internal, external, or a combination of both, and are required to receive higher credit ratings. The three most typical external credit enhancements are these:[9]

1. Cash collateral accounts: These are accounts funded when a series is issued and created by the securitization trust, which can be used to fund principal and/or interest on the certificates and other trust expenses when the excess spread turns negative.
2. Third-party letters of credit.
3. Collateral invested amounts (also known as CIAs): These consist of privately placed ownership interests in the securitization trust, and are subordinate in payment rights to all investor certificate holders.

The four most common internal credit enhancements are these:

1. Senior/subordinated certificates.
2. Spread accounts: If certain performance indicators fall below specific thresholds, any "excess spread" will be deposited into an account for the benefit of the CCR holders.
3. Excess finance charges: These are defined as the difference between the gross yield on the pool of securitized receivables and the cost of financing those receivables.
4. Overcollateralization.

35.8 CONCLUSION

This chapter has reviewed the most important characteristics of fixed-income derivatives and asset-backed securities. It reviewed two basic approaches for modeling the term structure of interest rates, and examined how it can be applied to the valuation of fixed-income instruments with embedded options. In addition, it described interest rate swaps, and discussed how investors can use them to hedge their interest rate risks.

Next, the chapter examined two types of asset-backed securities. Asset-backed securities can provide unique risk-return patterns to market participants. However, investors in these securities ought to dedicate time and resources to analyzing and performing due diligence on collateral, servicer, and structure. Investors can use the securities presented in this chapter either on a stand-alone basis or when searching for portfolio diversification.

The asset-backed securities market has recovered from the lows reached during the financial crisis of 2008–9, during which the market for many of these products (especially in the case of mortgage-backed securities) almost came to a halt in the United States, Europe, and Asia. New structured products are being launched. The attractiveness of asset-backed securities to issuers resides in the fact that their structures facilitate the removal of certain assets from their balance sheets. Nevertheless, the Dodd-Frank Act (2010) obliges issuers to preserve some of the credit risk of those securities. Asset-backed securities also allow issuers whose debt is below investment grade to issue investment-grade debt.

NOTES

1. This section is based mainly on Sundaresan (2002) and Fabozzi and Mann (2012).
2. See Bodie, Kane, and Marcus (2010) for a review.
3. The Cox, Ingersoll, and Ross (1985) model does not assume that the unbiased expectation hypothesis holds. This assumption is used here to simplify the discussion.
4. ISDA Research Note, "The Impact of Compression on the Interest Rate Derivative Market," July 2015.
5. Throughout this section, we ignore the precise day count that takes place in practice. In other words, it is assumed that there are 90 days in each quarter, 180 days in each six months, and 360 days in each year.
6. Alternatively, interest rate swaps can be valued as a portfolio of forward rate agreements (FRAs).
7. As stated, to keep the calculations simple, it is assumed that there are 90 days in each quarter.
8. In Exhibit 35.16, the column titled "Housing-Related" does not include mortgage-backed securities.
9. https://www.fdic.gov/regulations/examinations/credit_card_securitization/ch6.html.

REFERENCES

Altomonte, C., and P. Bussoli. 2014. "Asset-Backed Securities: The Key to Unlocking Europe's Credit Markets?" July 23. http://bruegel.org/2014/07/asset-backed-securities-the-key-to-unlocking-europes-credit-markets/.

Bodie, Z., A. Kane, and A. Marcus. 2010. *Investments*. 9th ed. Berkshire, UK: McGraw-Hill/Irwin.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53:385–407.

Culp, C. and J. Forrester. 2015. "Have Pre-Crisis Levels of Risk Returned in U.S. Structured Products? Evidence from U.S. Subprime Auto ABS, CLOs and Insurance-Linked Securities Markets." *Journal of Structured Finance* 21 (1): 10–44.

Cummins, J. D. 2008. "Cat Bonds and Other Risk-Linked Securities: State of the Market and Recent Developments." *Risk Management and Insurance Review* 11 (1): 23–47.

Davis, T. 2015. "Interest Rate Swap Valuation." White paper, FactSet Research Systems. https://www.factset.com/websitefiles/PDFs/whitepapers/interest_rate_swap_valuation.

Edesess, M. 2014. "Catastrophe Bonds: An Important New Financial Instrument." EDHEC-Risk Institute, July.

Fabozzi, F., and S. Mann. 2012. *The Handbook of Fixed Income Securities*. 8th ed. New York: McGraw-Hill Education.

Ferrara, P., and S. Ali. 2013. "Interest Rate Swaps: An Exposure Analysis." Society of Actuaries, July.

Heitfield, E., and T. Sabarwal. 2004. "What Drives Default and Prepayment on Subprime Auto Loans?" *Journal of Estate Finance and Economics* 29 (4): 457–77.

Ho, T., and S. Lee. 1986. "Term Structure Movements and Pricing Interest Rate Contingent Claims." *Journal of Finance* 41 (5): 1011–29.

Morrison, Joanne. 2015. "Cleared OTC Volume Increased in 2014." *Futures Industry Magazine*. March 9. https://fimag.fia.org/articles/cleared-otc-volume-increased-2014.

Securities Industry and Financial Markets Association. 2014. "Statistics on Asset Backed Securities." https://www.sifma.org/uploadedfiles/research/statistics/statisticsfiles/sf-us-abs-sifma.xls.

Smith, D. 2012. "A Teaching Note on Pricing and Valuing Interest Rate Swaps Using LIBOR and OIS Discounting." Boston University School of Management, June.

Sundaresan, S. 2002. *Fixed Income Markets and Their Derivatives*. 2nd ed. Cincinnati, OH: South-Western.

Vasicek, O. 1977. "An Equilibrium Characterisation of the Term Structure." *Journal of Financial Economics* 5 (2): 177–88.

CHAPTER 36

Structured Products II: Insurance-Linked Products and Hybrid Securities

This chapter continues the discussion of structured products. Many of these products are relatively new, and their participation in institutional portfolios is expected to continue to grow in the future. The first part of the chapter explores insurance-linked securities (ILS), a relatively new asset class. This section discusses two examples of ICS: catastrophe bonds and longevity-risk-related products.

The second part of the chapter discusses ways to invest in insurance risk. The important concept of mortality risk is analyzed, and the mechanics and investment attributes of life insurance settlements are discussed. Finally, the third part of the chapter covers mezzanine finance products. Here we discuss the investment vehicles available to access mezzanine finance: subordinated debt with step-up rates, subordinated debt with payment in kind (PIK) interest, subordinated debt with profit participation, subordinated debt with warrants, and convertible loans. We also comment on the use of these hybrid securities in project finance.

36.1 INSURANCE-LINKED SECURITIES

Insurance-linked securities (ILS) are tradable financial instruments whose value is affected by an insured loss event, such as a natural disaster, longevity risk, or life insurance mortality. ILS represent a convergence between capital and insurance markets. Institutional investors have increasingly regarded reinsurance as a new asset class, having invested around $50 billion in an array of insurance-linked securities over the past decade. ILS offer exposure to nonfinancial risks and thus are generally regarded as being uncorrelated with the general financial markets. Here we discuss two of the most important ILS groups of products: catastrophe bonds (a type of "non-life" ILS) and longevity-risk-related products (a type of "life" ILS).

36.2 OVERVIEW OF NON-LIFE ILS: CATASTROPHE BONDS

Catastrophe bonds (cat bonds) are risk-linked debt securities that are designed to transfer specific risks from issuers—typically insurance or reinsurance companies—to investors. These bonds represent the largest portion of the ILS market. These specific risks are usually those having to do with natural disasters, such as hurricanes and earthquakes in the United States, Europe, or Japan. Cat bonds are typically structured as private placements.

The occurrence of major catastrophic events in the early 1990s led to the development of the cat bond market in the second half of the decade. Those events were Hurricane Andrew in 1992 (Florida, United States, $17 billion of insured losses), the Northridge earthquake in 1994 (California, United States, $15 billion), and the Kobe earthquake in 1995 (Kobe, Japan, $3 billion). The large losses that these events imposed on insurance companies highlighted the need for supplementary risk capital throughout the industry, encouraging insurers and reinsurers to search for new sources of additional risk capital and new ways to spread the risk among a larger group of investors with the capacity to bear those risks. It was believed that this would allow insurance companies to continue offering protection against these types of events without the need to increase the insurance rate to such high levels that coverage would no longer be affordable. Furthermore, the low interest rate environment following the global financial crisis of 2008–9 encouraged institutional investors to search for investment opportunities in new high-yielding assets.

Cummins (2008) comments that the first successful cat bond was issued by Hannover Re in 1994. However, according to Sterge and Van der Stichele (2015), the first cat bond was issued by the St. Paul Companies, Georgetown Re, in December 1996. The cat bond industry grew exponentially from a few hundred million dollars in bonds outstanding at the end of the 1990s to around $13 billion by 2007. The growth in cat bond issuance came to a stop after the 2008–9 period because of the global financial and credit crisis and the bankruptcy of Lehman Brothers, which had sponsored four cat bonds, for which the market value of the investments in their special purpose vehicles (SPVs) was smaller than the fully redeemable value of the bonds. The market then resumed its growth, and by the end of 2014, the total amount of bonds outstanding reached a record $24 billion, of which $2.3 billion was in euros, $85 million was yen denominated, and the rest were U.S. dollar issues; see Sterge and Van der Stichele (2015). This represents around 8% of the traditional property catastrophe risk excess-of-loss reinsurance market. Exhibit 36.1 presents the evolution of total cat bond risk capital issued during the past decade.

Exhibit 36.2 shows the growth experienced by cat bond issuance and outstanding cat bonds during the past decade, as new sponsors and investors continued to access the market. The year 2014 marked a record in both risk capital issued ($8.03 billion) and risk capital outstanding ($22.87 billion).

36.2.1 Mechanics of Non-Life ILS: Catastrophe Bonds

Exhibit 36.3 describes how a cat bond works. First, a sponsor (insurer or reinsurer) enters a contract to transfer specific risks with an SPV or a securitization fund created specifically for that transaction. This securitization fund issues cat bonds to investors, and the money received from investors is invested in risk-free assets held in a collateral trust account, thus virtually eliminating credit risk. If no covered event takes place, investors receive variable interest (from the risk-free assets) plus a risk premium. However, if a catastrophic event takes place, the fund covers the losses of the protection buyer. When the bond matures, investors receive their principal minus

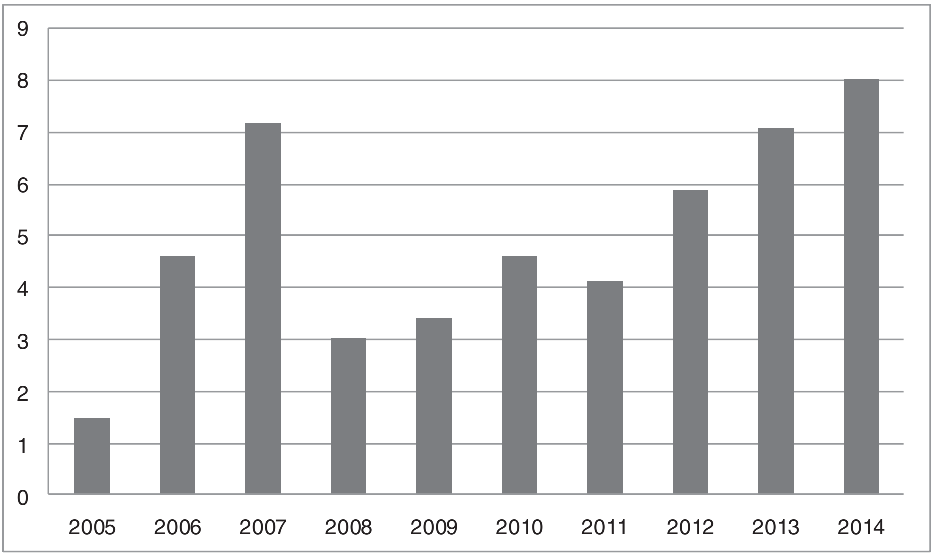


EXHIBIT 36.1 Risk-Based Capital Issues in the United States ($ billions)
Source: GC Securities.

any funds that would have been used to cover catastrophe losses during the life of the bond.

The capital supplied by investors when they buy cat bonds is used to compensate any covered losses from a catastrophic event as specified in the bond. If no catastrophic event occurs, investors receive all the coupon payments from the bond and the bond's principal. However, if a catastrophic loss does take place, investors will

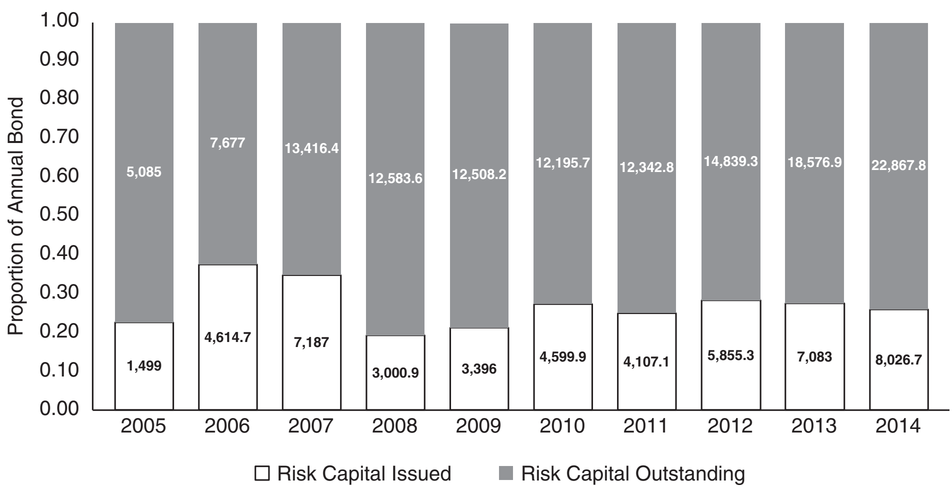


EXHIBIT 36.2 Cat Bonds Risk Capital Issued and Outstanding in the United States ($ millions)
Source: Guy Carpenter & Company, LLC (2015).

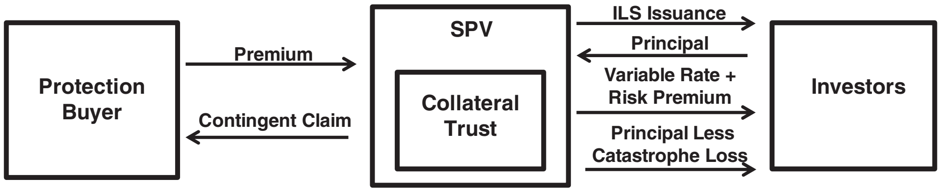


EXHIBIT 36.3 Diagram of a Cat Bond
Source: Based on a diagram presented in Weistroffer (2010).

usually forgo some of the coupon payments and suffer a partial or complete loss of principal. Furthermore, investors in cat bonds should be aware of the potential relative illiquidity of these securities.

Investors in cat bonds have received relatively high returns when no catastrophic event has occurred. Historically, such risks have had occurrence frequencies of 1 in 50 to 1 in 100 years, and investors have been paid a spread over a short-term benchmark that has fluctuated between 4% and 10% (Sterge and Van der Stichele 2015). The occurrence of a prespecified event triggers the loss of part of or the entire bond's principal, which is used to cover the issuer's indemnities.

Catastrophe bonds are attractive to investors because the occurrence of a trigger event is virtually uncorrelated with financial risks (e.g., market risk, credit risk, and interest rate risk). Maturities of cat bonds typically range between one and five years, with most bonds having a three-year maturity. Cat bonds are floating-rate bonds, and coupons are usually paid quarterly. Annual expected losses are modeled by three specialized firms and have historically ranged between 50 and 500 basis points (at issuance), with a size-weighted average of 160 basis points (Sterge and Van der Stichele 2015).

Cat bonds are generally used by insurers as a substitute to traditional catastrophe reinsurance. Reinsurance is insurance that is purchased by an insurance company from one or more other insurance companies, known as the reinsurer. The insurance company and the reinsurer enter into a reinsurance agreement, which details the conditions upon which the reinsurer would pay a share of the claims incurred by the insurance company.

The goal of a reinsurance program is to diversify the risk, which enables the insurance company to offer insurance at affordable rates to its customers, and to maintain its financial viability in the face of substantial insurance claims resulting from major and widespread natural disasters. A healthy reinsurance marketplace helps ensure that insurance companies can remain solvent because the risks and costs are spread out. By sponsoring cat bonds, both insurance and reinsurance companies can hedge their exposure to catastrophe risk.

36.2.2 Four Trigger Types of Cat Bonds

In a cat bond transaction, payment depends on the occurrence of a triggering event (Edesess 2014; NAIC 2012). There are four basic types of triggers: (1) indemnity triggers, which are based on the actual claims incurred by the sponsoring insurance company or companies; (2) industry loss triggers, which are based on an industry-wide index of claims; (3) parametric triggers, which are based on assumed claims from an actual physical event (such as the magnitude of an earthquake or the wind speed of a hurricane); and (4) modeled triggers, which are based on estimated claims generated by a computer model.

- Indemnity trigger. An **indemnity trigger** is a type of trigger that initiates principal reductions based on the level of actual excess claims paid by the issuer. As of August 2012, 37% of all natural catastrophe bonds outstanding had indemnity triggers (Swiss Re 2012).

  The **cat bond attachment point of the trigger** is a numerical value indicating the point at which at least a fraction of principal must be "attached" to cover claims. The **exhaustion point** is the level of claims loss at which the principal is "exhausted" and investors are not legally responsible for any additional claims. Related to attachment point is the **attachment probability**, which, typically based on historical information about natural disasters, indicates the estimated probability that the cat bond's attachment point will be reached.

  The indemnity trigger is advantageous for the issuer because it will have to pay claimants essentially the same amount the cat bond pays the issuer (i.e., there is no "basis risk"). On the other hand, the indemnity trigger is disadvantageous to the investor because it must wait until all claims are settled. More specifically, if there has been a triggering event (e.g., an earthquake), the investor may have to wait a long time after the bond matures to reclaim a portion or the entire principal. Furthermore, the indemnity trigger creates moral hazard because the issuer has the incentive to underwrite excessive risks—for example, homes built in areas where the risk of being struck by an earthquake is high—because actual losses are hedged. Because of this increased risk or moral hazard, investors usually require an extra return for investing in indemnity trigger transactions.
- Industry loss trigger. An **industry loss trigger** is a trigger in which principal reductions in the cat bond are based on index estimates made by an independent third party of the total industry losses due to the occurrence of an insured event. For example, consider the case of a triggering event that takes place when total industry losses exceed $8 billion. In this case, investors are accountable for the percentage of the industry that corresponds to the cat bond's issuer's share. As of August 2012, 40% of natural catastrophe bonds outstanding had industry loss triggers (Swiss Re 2012).

  In an industry loss trigger type, the issuer bears basis risk because the claims that it must pay may not correspond exactly to its share of the industry loss. Furthermore, the estimated total industry losses assessed by the independent third party and the actual total industry losses may not be exactly the same. The industry loss trigger type is more advantageous to the investor because claims are settled faster. This occurs once industry losses are estimated by the independent third party, which surveys the participants in the industry for their estimates. Furthermore, moral hazard is diminished given that the issuer bears some basis risk.
- Parametric trigger. A **parametric trigger** offers coverage when a certain threshold is surpassed based on previously specified natural parameters. According to Edesess (2014), a parametric trigger is based on the occurrence of a specific natural event, such as wind speed exceeding 120 km/hr (in a specified location), a category 5 hurricane, or an earthquake exceeding 7.0 on the Richter scale. While this poses basis risk to the issuer, it is advantageous to the investor because little

or no waiting time is required before settlement of the bond following a triggering event, resolution of losses is rapid and transparent, and the danger of moral hazard is low. However, investors in parametric trigger type cat bonds bear basis risk because there may be a mismatch between the actual loss incurred by the sponsor and the transaction recovery (Kusche 2013).

- Modeled trigger. In a **modeled trigger**, the coverage is based on claims generated by a computer model, developed by an independent modeling company. Catastrophe modeling software is used to estimate an exposure portfolio, which provides estimates of losses given various severities of a natural disaster. If a catastrophic event takes place, the event parameters are compared with the exposure portfolio, and the bond will be triggered if the modeled losses surpass a previously specified threshold. Therefore, given that involved parties do not have to deal with a company's actual claims, loss resolution after a triggering event can be faster, although the issuer preserves some basis risk.

36.2.3 Cat Bond Performance

The Swiss Re Cat Bond Total Return Index is one of the most widely used cat bond indices. This index tracks the total rate of return for all outstanding cat bonds denominated in U.S. dollars and is updated on a weekly basis each Friday. Swiss Re also calculates, among other indices, the following cat bond indices: global, U.S. wind, and California earthquake. Exhibit 36.4 shows the evolution of $100 invested in catastrophe bonds between 2002 and 2014 (Swiss Re Cat Bond Total Return Index), as well as the names of major catastrophic events that occurred during this period. It can be seen that the index has grown rather steadily since its inception, with only a few small and brief downturns in 2005, 2009, 2011, and 2013, each of which can be associated to the catastrophic events pointed out in the exhibit.

Exhibit 36.5 shows that between January 2002 and December 2014, an investment in cat bonds almost tripled its value in dollar terms, surpassing the returns

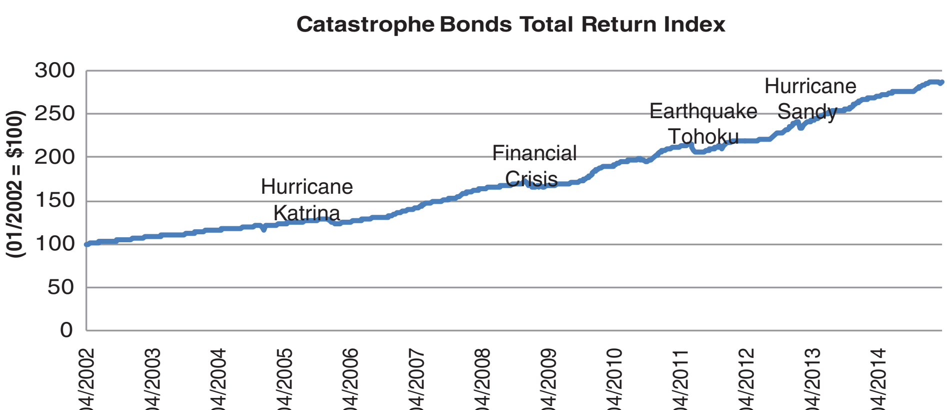


EXHIBIT 36.4 Catastrophe Bonds Total Return Index
Note: The graph shows the evolution of the Swiss Re Cat Bond Total Return Index.
Source: Bloomberg.

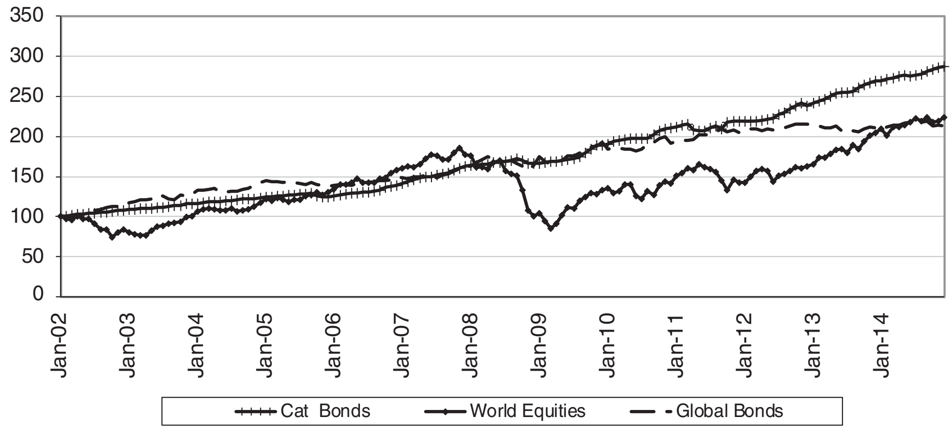


EXHIBIT 36.5 Cumulative Wealth Index (Cat Bonds, World Equities, and Global Bonds), 2002–2014
Source: Bloomberg and authors' calculations.

provided by world equities and global bonds, and at the same time exhibiting very mild fluctuations.[1] Furthermore, Exhibit 36.6 shows that cat bonds provided higher returns than world equities or global bonds and with a much smaller standard deviation (only 3.0% versus 15.7% and 5.9% for world equities and global bonds, respectively).[2] The Sharpe and Sortino ratios for cat bonds were much higher than those of any of the other four asset classes considered. Cat bonds had the smallest maximum drawdown of any of the five investments considered. In sum, cat bonds provided the highest returns and the lowest risk of the five asset classes considered here.

EXHIBIT 36.6 Statistical Summary of Returns

Index (January 2002–December 2014)	Cat Bonds	World Equities	Global Bonds	U.S. High Yield	Commodities
Annualized Arithmetic Mean	8.2%**	7.3%**	6.0%**	8.9%**	3.9%**
Annualized Standard Deviation	3.0%	15.7%	5.9%	10.0%	23.5%
Annualized Semivariance	0.0	1.2	0.2	0.4	2.7
Skewness	–0.8**	–0.8**	0.0	–1.2**	–0.6**
Kurtosis	5.4**	2.1**	0.9**	8.9**	1.5**
Sharpe Ratio	2.03	0.33	0.64	0.67	0.07
Sortino Ratio	0.372	0.047	0.093	0.112	0.010
Annualized Geometric Mean	8.1%	6.1%	5.8%	8.4%	1.1%
Maximum	3.6%	11.2%	6.6%	12.1%	19.7%
Minimum	–3.3%	–19.0%	–3.9%	–15.9%	–28.2%
Max Drawdown	–4.1%	–54.0%	–9.4%	–33.3%	–69.4%

* Significant at 90% confidence.
** Significant at 95% confidence.
Source: Bloomberg and authors' calculations.

Two recent papers, Kusche (2013) and Sterge and Van der Stichele (2015), find results similar to the ones reported here. Sterge and Van der Stichele (2015) also find that the correlation coefficients between the returns of cat bonds, U.S. bonds (Barclays U.S. Aggregate Bond Index), and stocks (S&P 500) are low, at only 0.2 and 0.4, respectively. Furthermore, between January 2001 and December 2014, a portfolio of issue-size-weighted cat bonds earned an average annual excess return of 6.6%, thus surpassing the 5.1% excess return that was expected on a modeled basis. The authors also note that both realized and modeled returns are comparable to the 5.9% historical excess return of U.S. large-cap stocks over the past five decades. However, the authors caution that, as of 2014, current market returns for cat bonds are much lower than historical returns.

Investing in cat bonds has some drawbacks. First, investors should bear in mind that these types of bonds are less liquid than most equities and corporate bonds issued by the largest firms. In addition, their payoff distribution is highly skewed and has significant tail risk to the downside. Jaeger, Müller, and Scherling (2010) argue that ILS returns are essentially driven by alternative beta. This means that they offer a positive expected return that should be considered compensation for the event risk assumed by investors. However, they argue that alpha-generating strategies based on security selection, market timing, and complexity arbitrage may exist. The first two sources are similar to what one finds in traditional equity and bond investments. However, complexity arbitrage arises because these securities are difficult to price and may trade in markets that are not entirely efficient. **Complexity arbitrage** is the process of attempting to earn short-term, very low-risk profits from pricing discrepancies attributable to highly complicated investment features. The authors caution that alpha strategies will be difficult to implement because many of them are costly to execute and constrained by their small size.

Finally, Kusche (2013) comments that many institutional investors that invest in cat bonds are not concerned about diversification within the catastrophe bond asset class because their allocation to cat bonds is often relatively small compared to their overall portfolio. Instead, these investors focus on the most attractive risk-return profiles among cat bonds and thus invest in, for example, U.S. wind-driven cat bonds, as these bonds have exhibited the highest risk-adjusted margins.

36.2.4 Establishing the Coupon Rate to Investors in Cat Bonds

This section examines a model that attempts to explain the market-clearing price of cat bonds.[3] First, we disaggregate the total coupon rate received by investors into two components: one that accounts for the time value of money (typically based on the LIBOR rate) and a second (the spread) that relates to the extra return demanded by investors for taking on the risk of suffering a potential cat loss:

$$\text{Total Coupon Rate to Investors} = \text{LIBOR} + \text{Spread} \tag{36.1}$$

The following equation is then used to estimate the spread for each exposure:

$$\text{Spread} = \text{Constant} + [\text{Loss Multiplier} \times \text{Expected Loss (\%)}] \tag{36.2}$$

where the loss multiplier parameter is a function of the uncertainty in the estimated expected loss. Since it is not possible to actually know the true value of the expected loss, one is forced to rely on cat modeling firms, which use computer software to produce estimates of the true expected loss. In its simplest form, expected loss is given by:

$$\text{Expected Loss }(\%) = \frac{\text{Probability of Event} \times \text{Annual Monetary Loss}}{\text{Principal}} \tag{36.3}$$

As an example of the use of this model, Bodoff and Gan (2009) use all years of data between 1998 and 2008 and fit the parameters of their model by inspecting results for U.S. wind cat bonds (related to losses arising from very high winds in Florida, Southeast United States, and/or Northeast United States), and obtain the following equation, which approximates the spread, when issued, of any cat bond that covers U.S. wind:

$$\text{Spread} = 3.33\% + [2.40 \times \text{Expected Loss }(\%)] \tag{36.4}$$

The intercept, 3.33%, and the slope, 2.40, are statistically significant variables. One can use expert judgment to improve the modeled spread by including some of the additional factors that affect the actual issuance spread (trigger type, market conditions, etc.).

APPLICATION 36.2.4

Suppose a three-year cat bond covering U.S. wind has just been issued. Based on the equation estimated by Bodoff and Gan (2009), with an intercept of 3.33% and a slope coefficient of 2.40, find the estimated spread assuming an expected loss of 1.40% per annum. Then, assuming three-year LIBOR is 1.7% per annum, calculate the total coupon rate (%) to investors for this bond.

First, recall that the spread is given by Equation 36.4. Therefore, we conclude that the spread should be equal to:

$$\text{Spread }(\%) = 3.33\% + (2.40 \times 1.4\%) = 6.69\%$$

Using Equation 36.1, the total coupon rate to investors should be equal to:

$$\text{Total Coupon Rate} = 1.7\% + 6.69\% = 8.39\%$$

Note that with both equations, any variable in the middle expression side could be computed if values of all the other variables were provided.

Finally, there are also other catastrophe-related securities. For example, cat derivatives are financial instruments indexed to catastrophic events. When one of these types of rare events occurs, these products pay a specified cash flow to their holders. The size of the payment is usually proportional to the characteristics of the event or to the losses suffered by the holder as a consequence of the catastrophic event. These types of financial securities may trade in organized exchanges or over the counter (OTC). Catastrophe derivatives are generally based on an index reflecting the severity of the catastrophic event.

Catastrophe bonds and other ILS may be subject to credit risk. However, even though this risk is supposed to be relatively small for these securities, the following five alternatives have been devised to mitigate and manage this risk:

1. The addition of an extra collateral account
2. A guarantee provided by a parent company
3. The issuance of a letter of credit by a bank or another financial institution
4. The use of credit derivatives (e.g., credit default swaps)
5. The use of credit insurance (i.e., insurance obtained through bond insurance firms)

36.3 LIFE ILS: LONGEVITY AND MORTALITY RISK–RELATED PRODUCTS

This section examines another class of insurance-linked securities, the life ILS. It discusses the concepts of longevity and mortality risk, two important and related risks. It then considers the investment attributes of life insurance settlements, a life insurance–linked security that has gained increasing interest in the past two decades, in spite of being controversial to some participants.

36.3.1 Longevity Risk

Longevity risk is any potential risk that arises from a higher than initially projected life expectancy of pensioners and policyholders. Life insurance companies and pension funds are, among other institutions, affected by longevity risk. Governments' fiscal balance sheets can also be severely affected by longevity risk, as we will explain.

Longevity risk has arisen mainly because people are living longer than expected (particularly since the 1990s), and as a result, payments associated with annuities purchased by retirees or pension funds have been extending longer than originally forecasted. For example, life expectancy at age 60 in advanced economies in Europe rose from 15 years in 1910 to 24 years a hundred years later, and it is expected to improve further, according to the International Monetary Fund (IMF 2012). This means that while in 1910 an annuity purchased by a 60-year-old retiree was expected to make 15 payments, the same annuity today is expected to make 24 payments. The most relevant measure of longevity risk is life expectancy at pensionable age.

According to the IMF (2012), over the past century, longevity predictions have underestimated observed mortalities by about three years. Furthermore, the exposure of pension funds and other institutions to unexpected improvements in life expectancy has been estimated to be more than $20 trillion worldwide (Biffis and Blake 2010).

Longevity risk represents a threat to financial stability for two reasons. One factor is the threats to the fiscal solvency of countries given that governments are exposed to longevity risk. A second factor is threats to the solvency of private corporations and financial institutions that are vulnerable to longevity risk, mostly through their defined benefit pension plans.

The exposure of governments to longevity risk is large, and thus it increases the already large costs of aging populations. This occurs in the midst of weak fiscal balance sheets by the governments of many countries. According to a report by the IMF (2012), if everyone lives three years longer than expected, then the present value of the extra living expenses during those additional years of life would represent between 25% and 50% of 2010 global gross domestic product (GDP).

Governments are exposed to longevity risk in three ways: (1) through their liabilities in defined benefit public pension plans, (2) through social security programs, and (3) through being regarded as the "holders of last resort" of the longevity risk borne by financial institutions and individuals.

Longevity improvements around the world are attributable to declining infant mortality rates, improved living conditions, and better medical care. Longevity risk can be severe for plans such as annuities and defined benefit pension plans because even modest adjustments in life expectancy can cause serious solvency problems. For pensioners on defined contribution plans, the problem is complicated by the fact that these schemes typically do not offer longevity risk protection.

Longevity risk has usually been estimated by analyzing historical mortality data and then forecasting based on past trends. Unfortunately, projections can vary substantially, with only modest variations in inputs such as, for example, the time period of the data used. Moreover, random noise often generates outlier paths that are medically improbable. On this, Coburn and Nakada (2012) comment:

> *Research from the fields of medical science and demography... [is] used to understand the drivers of historical mortality change and to formulate models of how... medical advance may drive further change. [This] allows the construction of future scenarios that can be used to explore... scenarios of tail risk. (93)*

36.3.2 Hedging Longevity Risk

Longevity risk can be hedged using index-based and indemnity-based longevity swap contracts (OSFI 2014). A **longevity swap contract** takes place when a pension plan administrator agrees to make fixed payments to a counterparty based on certain mortality assumptions, while the counterparty agrees to make floating payments based on either the pension plan's actual mortalities (indemnity-based contract) or an agreed-upon mortality index (index-based contract). Thus, these contracts allow the counterparty to assume the longevity risk for a price. Exhibits 36.7 and 36.8 illustrate the functioning of each contract. Note that the margin included in the fixed cash flows stands for the risk premium demanded by the counterparty to assume longevity risk and expenses.

In the case of indemnity-based longevity contracts, if a pension plan's beneficiaries live longer than was originally assumed, the pension plan will receive higher

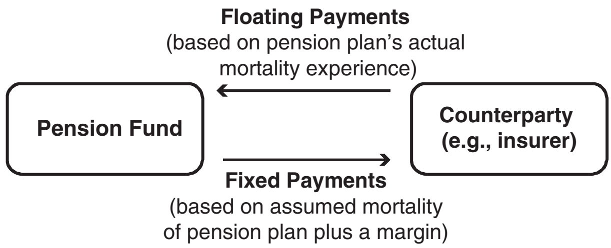


EXHIBIT 36.7 Indemnity-Based Longevity Contract
Source: Based on OSFI (2014).

payments from the counterparty, thus helping to offset the plan's ex post higher pension costs. In the case of index-based contracts, if there is an increase in general longevity (as measured by the respective index), this will cause higher payments from the counterparty to the plan.

Index-based longevity swaps have the advantage that they are readily available and much more marketable than indemnity longevity swaps. Index-based longevity swaps can also bring in counterparties that do not want to hold specific longevity risk.[4]

Pension plans that use these contracts to hedge longevity risk are exposed to four risks:[5]

1. Counterparty risk. This is the risk that the counterparty to the contract will not fulfill its obligations.
2. Rollover risk. This risk arises when pension plans agree to enter into hedging contracts for a shorter period of time than the liabilities they are aiming to cover.
3. Basis risk. This risk arises when index-based longevity risk hedging contracts are used, as there is a chance that the mortality experience of the pension plan could diverge from that of the index associated to the respective contract.
4. Legal risk. This risk arises because longevity risk hedging contracts are not exchange traded, and thus it is highly advisable that the parties agreeing on a contract scrutinize its terms carefully.

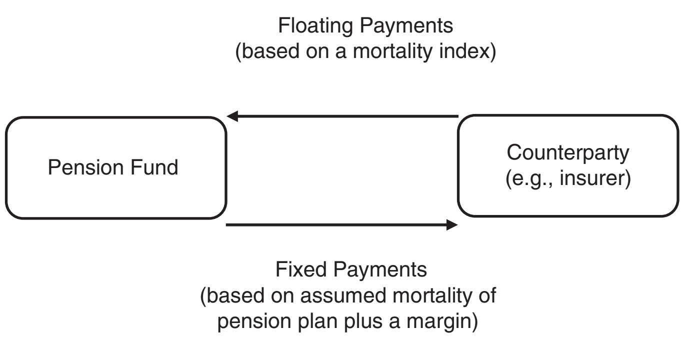


EXHIBIT 36.8 Index-Based Longevity Contract
Source: Based on OSFI (2014).

36.3.3 Mortality Risk[6]

Mortality risk is the risk of a person (or group of individuals) passing away sooner than expected. Such an event could create financial distress at the individual level if the deceased is the family wage earner and passes away without life insurance. This risk is also borne by companies writing life insurance if a large number of sooner-than-expected mortalities occur. It can be seen that longevity risk is the flip side of mortality risk.

A related concept is that of extreme mortality risk. **Extreme mortality risk** arises because of the threat of very high mortality rates due to natural disasters, pandemics, and terrorist attacks. Reinsurance of extreme mortality risk has become very expensive after the terrorist attacks of September 11, 2001.

Swiss Re, the world's largest life and health reinsurance company, was the first company to design a financial security in order to transfer extreme mortality risk directly to the capital markets. This happened in 2003 through the Vita transaction. In this transaction, Swiss Re structured a catastrophic mortality bond to reduce the risk to the firm of an adverse mortality experience. The trigger was established to be a weighted average of the overall population mortality rates in five countries (France, Italy, Switzerland, the United Kingdom, and the United States) to mitigate the exposure of Swiss Re to a catastrophic mortality event in those five countries.

An SPV, Vita Capital Ltd., was created for the securitization. The SPV entered into an agreement with Swiss Re to offer a call option on the SPV assets, in exchange for a premium paid by Swiss Re. At the same time, the SPV issued a bond to investors. Returns provided by the collateral account were swapped with a highly rated counterparty in exchange for a LIBOR-based interest rate.

The transaction works as follows. Swiss Re will receive payments arising from the reduction of principal repayment to bondholders. This will happen if the index value goes above a threshold of 130% of the base value (known as the attachment point), and then increases proportionally until it reaches a threshold of 150% (known as the exhaustion point), at which point Swiss Re receives the full amount and investors receive no principal repayment.

Subsequent issuances of extreme (catastrophic) mortality bonds have included tranches. For example, the transaction may include two tranches with different attachment and exhaustion points. Another advance has been made in the use of a credit wrap to improve the credit rating of a specific tranche. A credit wrap is a credit enhancement in which an insurance company guarantees the payment of interest and principal of a specific debt in exchange for an insurance premium. Wrapped bonds yield lower spreads because of the risk reduction that they achieve.

The main risk in catastrophic mortality bonds is assumed to come from a pandemic or a similar disease that may severely affect children and the elderly. These types of extreme and rare events are very difficult for insurance companies to model.

According to Krutov (2010), the following factors can affect mortality rates: catastrophic events, random fluctuations, misestimation of mortality trends, and miscalculation of claim levels and data issues. Furthermore, extreme mortality securitization entails independent modeling of the following three main elements of mortality rates: (1) baseline mortality (reflecting statistical fluctuations), (2) terrorism element (effect of terrorist acts on mortality rates), and (3) pandemic component (effect of major epidemics of severe infectious diseases on mortality rates). Pandemics are the main cause of potential jumps in mortality rates. Most life insurance policies do not include war as a covered cause of death.

Catastrophic mortality bonds may offer lower diversification benefits than those provided by non-life cat bonds. The reason is that an event such as a pandemic could lead to disruptions in many markets and therefore negatively impact most financial securities, increasing their correlations.

36.3.4 Overview of Life Insurance Settlements

Life insurance settlements, or life settlements, consist of the sale of a life insurance policy (or the transfer, bequest, or assignment of a life insurance policy or the benefits of such policies) by its owner to a third party. The buyer of the policy becomes its beneficiary and assumes the payment of the premiums. Life insurance policies in life settlements have to be from policyholders who are not terminally ill (even if they are sick and elderly). There is a dispute in some U.S. states and in other countries about the legality of life insurance settlements.

The cash **surrender value** of a policy is the price at which the insurance company will buy back its commitments under the contract. If the price offered to the owner of a life insurance policy by a third party is greater than the cash surrender value of the policy, it could be beneficial to the policyholder to sell the policy to a third party. From the point of view of an investor, the transaction could be valuable if the discounted value of the future benefits of the life insurance policy exceed the combined values of the discounted premium payments to be made and the cost to acquire the policy. There are rational economic reasons for a policyholder to sell a life insurance policy. The investor who purchases the insurance policy may still benefit from the purchased policy if the price paid is less than the net present value of all the cash flows (premiums and life insurance benefits) associated with the contract. In other words, both parties benefit if:

$$\text{Surrender Value} < \text{Purchase Price} < \text{NPV of Cash Flows} \tag{36.5}$$

Of course, the beneficiaries of the insurance policy are giving up potential benefits equal to the difference between the NPV of the cash flows and the face value of the policy purchase price. However, the beneficiaries gain immediate access to cash and do not run the risk of being unable to pay the premiums.

Many life insurance policies are structured in a way whereby the premium payments remain level even though the rate of mortality increases over time. Effectively, in the beginning, the premiums paid on such a policy are higher than necessary for the expected level of claims. After a certain period, however, the situation reverses and the premiums no longer cover claims and other expenses as mortality goes up with age. The policy is still profitable to the life insurance company because the overpayment in the beginning more than offsets the underpayment toward the end of the policy term. Reserves that have been built up from the beginning are used to pay claims, most of which come later. This simplified example further demonstrates how an insurance policy could have monetary value to the policyholder who has been paying premiums for several years. On an expected basis, the net present value of the future premiums could be lower than the net present value of the death benefit, often by a significant amount. The difference is even greater for a policyholder whose health condition has significantly deteriorated since the initial underwriting and whose mortality rate has thus increased beyond the expected value. The value of such policies to potential investors has correspondingly gone up (Krutov 2010, 264–65).

36.3.5 Modeling Life Insurance Settlements

The typical life insurance settlement can be represented as a bond having negative, fixed-coupon payments (the annual premiums), a known principal (the face value of the policy), and an unknown duration (the policyholder's life expectancy). For example, consider the case of a policyholder with a life expectancy of 10 years and a life insurance policy with a face value of \$1,000,000. The annual premium on the policy is 3%. Assuming a discount rate of 10% and ignoring commissions, fees, and taxes, the net present value of the policy to the policyholder can be calculated. The policyholder has to make annual payments of \$30,000 on the policy (i.e., \$1,000,000 × 3%) during each of the 10 years. The net present value of the policy to the policyholder is then:

$$\mathrm{NPV} = \sum_{t=0}^{10} \frac{-30{,}000}{(1+0.1)^t} + \frac{1{,}000{,}000}{(1+0.1)^{10}} = 201{,}206$$

If the surrender value of the insurance policy is less than NPV (\$201,206), then there will be a price at which both the policyholder and the investor could benefit from the transaction. Note that the calculation is similar to that of a bond with a negative coupon payment of \$30,000 and a positive face value of \$1,000,000.

APPLICATION 36.3.5

A life insurance owner has a life expectancy of five years and a life insurance policy with a face value and death benefit of \$500,000. Her annual premium on the policy is 4%, and its cash surrender value is \$200,000. Find the excess of the NPV of the policy to its cash surrender value if the market discount rate is 8%.

The present value of five years of \$20,000 premium payments represents a liability to the policy owner of –\$79,854. The present value of the death benefit in five years is +\$340,292. The net present value to the policy owner is \$260,437, which exceeds the cash surrender value by \$60,437.

36.3.6 Overview of Viatical Settlements

A **viatical settlement** is a transaction in which a sick policyholder sells his or her life insurance policy at a discount to its face value. The buyer obtains the face value of the policy when the original policyholder passes away. It is sometimes difficult to distinguish viatical settlements from life settlements. However, the transaction will be regarded as a viatical settlement if the life expectancy of an insurance policyholder is less than two years.

The viatical settlement industry developed in the late 1980s and early 1990s when many people diagnosed with AIDS had purchased life insurance policies before they were diagnosed, and these settlements offered the policyholders lump sums of cash that helped finance the expensive medical costs related to the illness. Subsequently, in the 1990s, many states in the United States adopted viatical settlement laws, which contributed to the growth of a secondary market for life insurance policies.

Viatical settlements are tightly regulated in many U.S. states. Some regard investors in viatical settlements as providers of an important public service, because they allow a person with a terminal disease to obtain funds when they are most needed (e.g., to pay for better medical care).

In addition to the United States, other countries have developed markets for life insurance policies. For example, in Germany and the United Kingdom, traded endowment policies (TEPs) allow investors who hold them until maturity to receive the terminal bonus payments of these policies. The life settlement market is essentially institutional. The market for life insurance policies is expected to continue to grow as investors and policyholders gain knowledge about this transaction, and also as baby boomers age. Unfortunately, many policyholders do not know that they have the option to sell unaffordable or unnecessary life insurance policies, and let their policies lapse. Furthermore, a life insurance policy could be used as collateral for a loan if it is placed in a trust.

36.3.7 Investment Benefits and Risks of Viatical Settlements

From an investor point of view, life settlements offer the potential advantage of having low correlations with traditional investments. However, and similar to the case of other fixed-income assets, the value of life settlements can be affected by interest rate changes. Life insurance policies and other longevity-based financial securities tend to be uncorrelated with most other assets, making them good candidates for enhancing portfolio diversification.

Unfortunately, and similar to the case of catastrophic mortality bonds, data on the investment performance of life settlements is scarce. Furthermore, life settlements are very illiquid investments. These two realities complicate the estimation of the value of investments in life settlements and the calculation of net asset values to be reported to investors (e.g., in the case of hedge funds investing in life settlements). Finally, another potential drawback of investing in life settlements is that they often involve substantial legal risks. Legal due diligence is thus an essential component of the life settlement investment process.

The existence of a universal exchange would make possible an open market with price transparency, facilitating the management of longevity and mortality risks. Synthetic products, mainly indices, have been launched in response to the growth of longevity and mortality markets. Indices have been developed based on a pool of actual lives. For example, Goldman Sachs created the QXX Index in 2007, a synthetic product that included almost 50,000 actual lives. The underlying pool to the QXX Index was provided by American Viatical Services. However, this longevity index proved to have a short life, as it was discontinued two years later on the grounds of investors losing interest in the product after the financial crisis, and also because of the “headline risk” that this kind of product may have brought to Goldman Sachs.

As previously commented, the lack of publicly available data on life settlement transactions makes it difficult to estimate the risk and return of these investments. Braun, Gatzert, and Schmeiser (2011) obtain an indirect estimation of the investment attributes of this asset class by analyzing the net asset value (NAV) performance of 17 open-end funds that were exclusively dedicated to investing in U.S. life insurance settlements between December 2003 and June 2010. During this six-and-a-half-year period, life settlement funds recorded a cumulative return of 37.30%. Stocks (S&P 500) provided a negative cumulative return of 2.60% during the same period. However, hedge funds recorded a higher cumulative return of 45.90%, and government bonds recorded a similar 37.38% return over this period. Furthermore, an investment in life settlements via open-end funds suffered only a relatively moderate drawdown during the financial crisis of 2007–9.

However, results from this study should be taken with caution. The calculation of the NAVs of these funds is subject to potentially severe valuation and liquidity risks. Other risks that affect an investment in life insurance settlements include longevity risk, policy availability risk, credit risk, operational risk, and changes in tax legislation and regulation.

In a more recent study, Januário and Naik (2014) examine settlement transactions by original policy owners using a data set containing all information in relation to 9,000 policies with aggregate net death benefits of $24 billion. These policies had been purchased in the secondary market by Coventry First from original policy owners between January 2001 and December 2011. The authors find that

> *by selling their policies in the secondary market, policy owners received more than four times the . . . cash surrender value they would have received had they surrendered their policies to their respective life insurance companies. . . . The [annualized] internal rate of return on the life settlements in our sample is 12.5%. (3)*

36.4 HYBRID PRODUCTS: MEZZANINE DEBT

Mezzanine finance, also known as mezzanine debt (MD), has already been covered in the CAIA curriculum in the context of private equity investment. Here, we center the discussion on the characteristics of the main investment vehicles available to mezzanine financiers to provide funds to firms.

Companies use mezzanine finance most commonly when they have entered an expansion phase, are generating positive cash flows, and are already profitable. Mezzanine debt entails an obligation to make interest payments and ultimately to make additional payments based on the performance of the firm. Mezzanine financiers invest with the expectation that debt payments will be honored and that they will exit the company without acquiring control or participating in its management.

Mezzanine debt can be compared to venture capital (VC). In the latter, investors are willing to finance companies that have negative cash flows (and are expected to remain negative for a certain time) in exchange for high rates of return. Furthermore, venture capitalists perform an active role in supervising the development of the firms in which they invest. The following section discusses the main characteristics of mezzanine finance.

36.4.1 Overview of Mezzanine Finance

Mezzanine finance is a hybrid of debt and equity financing typically used to fund expansion projects, recapitalizations, acquisitions, and management and leveraged buyouts. MD offers a risk-return profile that lies between debt and equity and is used as an alternative to equity to enhance the financial leverage of transactions. MD is normally used when a company has reached its senior debt capacity levels but its cash flows are substantial enough to withstand the costs of servicing newly issued long-term debt. As a result, and when well structured, MD reduces the amount of equity required in a business, thus lowering a company's weighted average cost of capital and enhancing stockholders' returns.

Mezzanine investing gained visibility during the global financial crisis of 2008–9. The resulting shortage of capital and the dislocation in the credit markets led to a surge in transactions. The prominence of mezzanine financing during the crisis, in turn, created further awareness on the fundraising part of the market, and many financial institutions went into the mezzanine lending area or raised extra capital for existing mezzanine funds (Robinson, Fert, and Brod 2011).

Exhibit 36.9 shows a comparison of mezzanine finance with senior debt and equity.

MD consists mostly of subordinated, unsecured debt and convertible, subordinated debt. In those cases in which MD is issued as senior debt at the same rank

EXHIBIT 36.9 Comparison of Mezzanine Finance with Other Financing Techniques

	Senior Debt	Mezzanine	Equity
Economic Perspective	Debt	Equity	Equity
Legal Perspective	Debt	Equity	Equity
Ranking	Senior	Contractually subordinated	Junior
Taxation	Debt interest deductible	Debt interest deductible	Tax on capital
Covenants	Comprehensive restrictions	Tracks senior, but looser	None
Security	Yes—1st ranking	Yes—2nd ranking	No
Investor's Involvement in Management	No direct involvement	Moderate involvement; board seats	Direct involvement
Purpose	Contractually specified	Not specified	Not specified
Term	4–5 years	5–10 years	Open ended
Interest Costs	Cost of funds + 255–350 basis points	150–300 basis points above senior	None
Repayment	Amortizing from cash flow	Bullet* upon exit or at maturity	None
Warrants	None	Almost always	None
Total Expected Return	5–10%	10–15%	>15%

* The payment for the principal is not made over the life of the loan, but rather as a lump-sum payment at exit or maturity.

Source: Center for Entrepreneurship, SMEs, and Local Development (2013); adapted from Credit Suisse (2006).

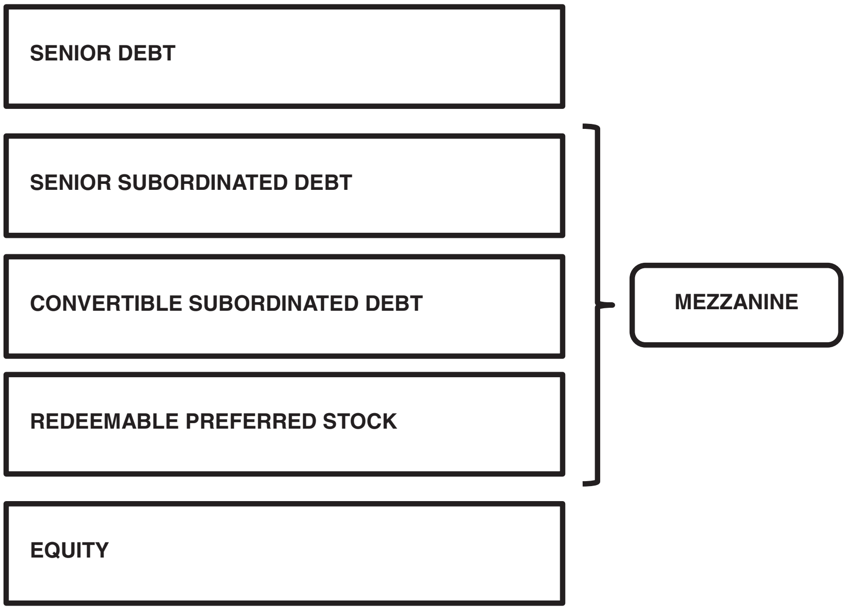


EXHIBIT 36.10 Senior Debt, Mezzanine Debt, and Equity
Source: Based on Silbernagel and Vaitkunas (2012) and Fitch Ratings.

with other debts issued by the firm, the remaining senior debt is secured. MD also takes the form of redeemable preferred stock. The investor base in mezzanine finance includes pension funds, hedge funds, high-net-worth individuals, family offices, leveraged public funds, and banks. Mezzanine financiers follow a buy-and-hold strategy. The majority of MD includes mandatory prepayment/redemption provisions, which become binding when certain events occur (e.g., asset sales).

MD is senior in priority of payments to common stockholders but subordinate to senior debt (see Exhibit 36.10). According to Silbernagel and Vaitkunas (2012), the funding gap between senior debt and equity usually occurs for one of four reasons:

1. Accounts receivable, inventories, and fixed assets have been discounted at higher rates than in the past, anticipating that their values will not be realized.
2. Senior debt lenders are unwilling to lend when intangible assets or goodwill is used as collateral.
3. Senior debt lenders may have reached a limit in their exposure to a certain company or industry.
4. Equity may simply be unavailable, insufficient, greatly dilutive, or too expensive.

36.4.2 Benefits and Disadvantages of Mezzanine Debt to Issuers and Investors

Mezzanine finance offers many advantages to firms with financing needs. First, it reinforces economic equity capital without the need to give up ownership rights or dilute equity. Second, MD remedies financial needs and facilitates capital available to undertake projects. Third, interest payments are tax deductible. Fourth, MD strengthens a company's cash position, thus facilitating the future financing of the firm using other sources of capital. And fifth, MD provides greater entrepreneurial autonomy for the company.

On the negative side, MD is more expensive than the typical loan financing, the capital is provided only for a limited time (compared to the case of equity capital), and the issuance of mezzanine debt requires that the issuer implement stricter transparency requirements.

Mezzanine financing has a debt component and an equity component. The mezzanine investor benefits when a company succeeds and does not cause a major dilution of existing equity. The debt part provides exposure to downside risk.[7] It is usually constituted by loans maturing in five or more years, with semiannual or quarterly fixed-rate coupons that are typically higher than senior debt and offer the possibility of rolled-up interest. Other usual terms include security and covenant protection. The equity part offers the potential for upside participation through equity kickers (warrants, conversion features, profit participation, or additional interest related to financial variables), exit through sale to a strategic investor or an initial public offering (IPO), or through redemption features, put options, and/or prepayment premiums (see Exhibit 36.11).

36.4.3 Terms and Yields of Mezzanine Debt

Mezzanine investors generally require a higher internal rate of return (IRR) on their investment than high-yield or bank loan investors. However, whereas VC investors target an IRR of more than 25%, mezzanine investors generally target an IRR that ranges between 15% and 25% (Silbernagel and Vaitkunas 2012). The targeted IRR is expected to be realized by a combination of fees, the interest rate, and the equity portion of the financing.

Covenants used in MD are typically based on high-yield type covenants or bank facility covenant packages. Key negative covenants in mezzanine finance may consist of limitations on asset sales, debt issuance, control changes, restricted payments, affiliate transactions, and liens. Affirmative covenants may require maintenance of insurance, financial reporting, and compliance with certain regulations.

Specialist mezzanine investment funds and other mezzanine lenders look for a target rate of return, which can be earned through two basic components: current payments and deferred payments. These payment streams can be further divided as follows:[8]

- Current payments (can be paid monthly, quarterly, or annually):
  - Cash interest: This is a periodic fixed or floating cash payment calculated as a percentage of the outstanding balance of the mezzanine financing.
  - Principal: This consists of scheduled repayments, a fraction of which may be deferred until maturity.
  - Royalties: These are variable payments based on a prespecified formula often associated with revenue, EBITDA, gross margin, or net income.
- Deferred payments (can be paid upon maturity of the mezzanine financing or later):
  - Payment in kind (PIK) interest: In a **payment in kind (PIK) interest** agreement, periodic interest is paid by increasing the total principal via capitalization of the interest payments due rather than in cash.

EXHIBIT 36.11 Summary of Mezzanine Debt Products and Attributes

Debt with Warrants	Convertible Loan	Participating Loans	Debt w/ Step-Up or Deferred Interest	Convertible Preferred Equity
Typically subordinated debt with detached equity warrants Principal is repaid after senior debt is largely or fully repaid Cash coupon Equity warrants can have zero or very low exercise price, and represent least dilutive minority stake in issuer	Typically subordinated Principal repaid at maturity unless conversion occurs, in which case debt swaps into equity at pre-agreed formula Inherently more dilutive than debt w/ warrants Most applicable on eve of IPO or when seed capital PE investor seeks max downside protection	Alternative form of mezzanine Base interest rate plus a performance-linked spread Interest rate linked to net profit, EBITDA, or sometimes sales; easy to game No equity Cumbersome to structure in practice	Subordinated debt with large amount of interest deferred (PIK often at final maturity) No equity participation, but potential for high IRR nonetheless Principal repayment is typically back ended Does not work well for longer-term finance (cumulative effect of deferred interest becomes punitive)	Similar to convertible debt, except that default on payment does not accelerate other debt and cannot force issues into bankruptcy Essentially a form of senior equity Widely used in start-up financings; basically equity
Direct	**Direct**	**Indirect**	**Subordinated debt**	**Equity**
Senior debt with equity participation				

Source: Bussman (2011).

- Equity ownership: Mezzanine capital will often include an equity stake in the form of attached warrants, a debt-for-shares conversion feature, or common shares of the company.
- Bonus payment: A **bonus payment** is an extra negotiated payment in a debt contract, which can be fixed or variable. In the case of the latter, payments are usually calculated based on the change in the value of the company over the length of time of the mezzanine financing or as a proxy for the business value.

36.4.4 Mezzanine Products

This section presents an overview of the mechanics of the basic mezzanine products.[9] These include subordinated debt with step-up rates, subordinated debt with PIK interest, subordinated debt with profit participation, subordinated debt with warrants, and convertible loans. Subordinated debt (also known as junior debt) is the most common form of mezzanine financing.

As will be seen, some of these products provide exposure to equity through an actual stock participation (these include subordinated debt with warrants, and convertible loans). Other products have exposure to the equitable uptick that the firm might experience (these include subordinated debt with step-up rates, subordinated debt with PIK interest, and subordinated debt with profit participation).

36.4.4.1 Subordinated Debt with Step-Up Rates This MD product is used in cases in which a firm cannot take on more debt with a fixed-cost scheme, because the current levels of senior and subordinated debt are exhausting the current cash flows. Interest rates in subordinated debt with step-up rates increase as the debt matures; therefore, it represents a supplementary level of subordinated debt that can be adapted to the firm's projected cash flows following a time-based or a criteria-based schedule. More often, a hybrid model that combines the two is used. Let us illustrate these concepts using the following examples.

Consider firm XYZ, which cannot take on more debt with a fixed-compensation mechanism due to the high levels of senior debt it already has on its balance sheet. The firm is offered subordinated debt with the time-based step-up rate mechanism depicted in Exhibit 36.12, in which interest payments increase toward the end of the life of the loan (a time at which, presumably, the company will have both created additional cash flows to service the debt and paid off the outstanding senior debt).

A problem with the time-based step-up rate is that there is no guarantee that the company will have experienced sales and margins high enough to honor its debt in the years in which interest rates increase. An alternative consists of designing a mechanism in which interest rates increase only if certain financial parameters have been met, such as in the next example.

Suppose that firm XYZ will pay higher interest rates in future years only if the interest coverage ratio (ICR) is greater than a certain value (see Exhibit 36.13).

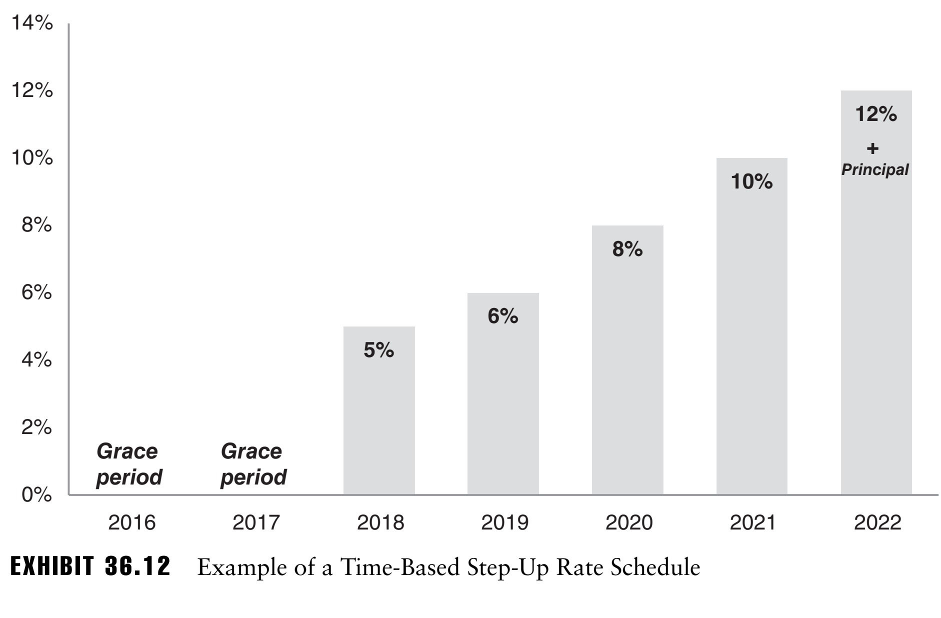


EXHIBIT 36.12 Example of a Time-Based Step-Up Rate Schedule

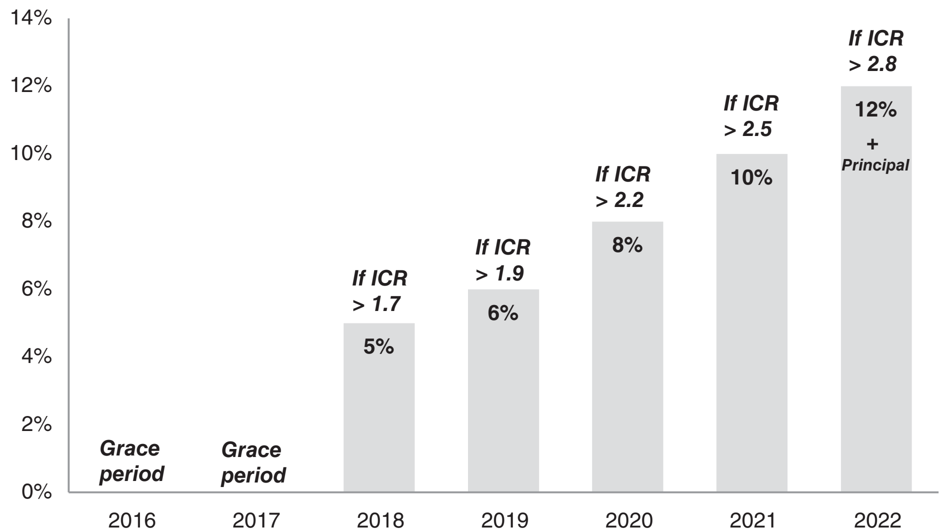


EXHIBIT 36.13 Example of a Criteria-Based Step-Up Rate Schedule

(*Note:* Use of the ICR is just one example; any agreed-upon criterion could be used.) For example, the firm will pay an interest rate of 8% in 2020 if the ICR exceeds 2.2 that year.

The time-based and criteria-based step-up rate mechanisms are often too rigid. In practice, market participants use a hybrid model step-up schedule in which a certain level of fixed interest rates are paid each year, independent of firm performance (this would be the time-based step-up rate part), and additional interest is paid depending on firm performance and based on an agreed criterion (this would be the criteria-based step-up rate part). The following case illustrates an example of this hybrid model.

Consider the case in which firm XYZ is charged a 5% interest rate on the loan independent of firm performance in 2018, and an extra 2% interest rate if the interest coverage ratio exceeds 1.7 (see Exhibit 36.14). In the following year (2019), the firm will be charged 6%, independent of firm performance, and an extra 2% if the ICR is greater than 1.9. The payment schedule for subsequent years is depicted in the exhibit.

36.4.4.2 Subordinated Debt with PIK Interest Subordinated debt with PIK interest is a type of obligation that does not provide any cash flows (interest or principal repayment) from the borrower to the lender between the drawdown date and the maturity of the loan or the refinance date. This means that both interest and principal payments become due in one balloon (bullet) payment when the debt matures.

Consider a five-year, £1,000,000 non-amortizing loan at 12% PIK annual compounding interest rate. The total PIK interest to be paid by the borrower can be calculated as follows. The PIK interest owed for the first year would be £120,000 (i.e., £1,000,000 × 0.12), which is paid in a security and is added to the principal amount of the debt, increasing the total amount owed to £1,120,000 at the end of

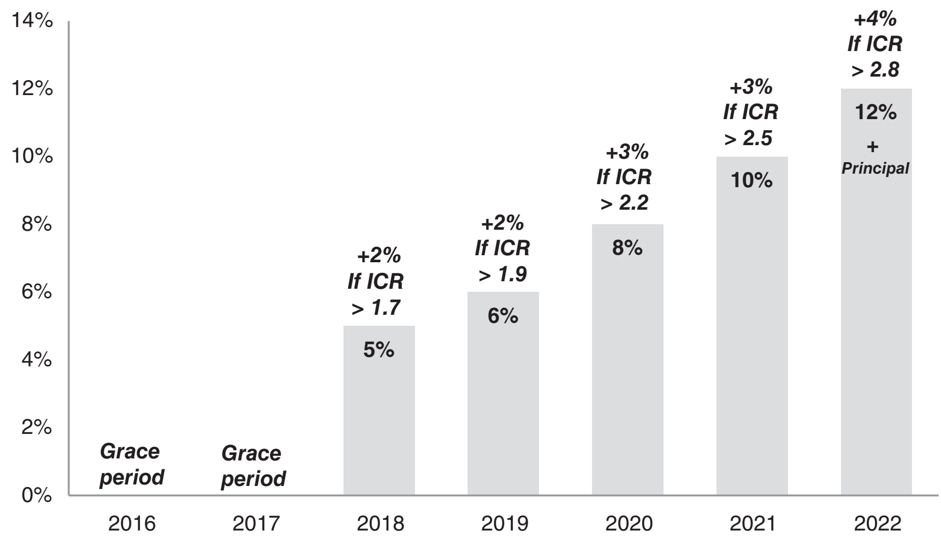


EXHIBIT 36.14 Example of a Hybrid Model Step-Up Rate Schedule

EXHIBIT 36.15 PIK Interest Example

Year	Beginning Debt Balance	PIK Interest	Ending Debt Balance
1	£1,000,000	£120,000	£1,120,000
2	£1,120,000	£134,400	£1,254,400
3	£1,254,400	£150,528	£1,404,928
4	£1,404,928	£168,591	£1,573,519
5	£1,573,519	£188,822	£1,762,342
Total PIK Interest		£762,342	

the first year. During the second year, the 12% annual rate is applied against the new principal balance, resulting in a total amount of £134,400 (i.e., £1,120,000 million × 0.12). The same process continues until the fifth year (see Exhibit 36.15).

The total PIK interest, which will be paid in the fifth year, amounts to £762,342 (rounded).

APPLICATION 36.4.4.2

A five-year bond with an initial principal amount of $2,000,000 is a non-amortizing loan with a 10% PIK annual compounding interest rate. Calculate all the cash payments that investors will receive, assuming that there is no default.

Without amortization and with the PIK feature, there will be no cash interest payments prior to maturity. The principal of the bond will increase at the

10% PIK annual compounding interest rate through the entire maturity of the bond (five years). The \$2,000,000 principal amount, increased with compounding for five years at 10%, grows to \$3,221,020, which is the total cash due when the bond matures in five years. The value may be found using the same math as is used to compute future values.

Both step-up rates and compounded interest payments are tax deductible in most countries. PIK loans are either unsecured or deeply subordinated obligations and entail a high-risk profile. Interest on PIK loans provides the lender with three sources of cash flow: an arrangement and/or a commitment fee, the (accrued or compounded) interest, and sometimes a ticking fee—a fee that is paid by the borrower to the lender to account for the time lag between the commitment on a loan and the actual disbursement.

PIK loans usually have maturity dates of seven to eight years, although terms can be as long as 10 years (with a minimum term of five years). Refinancing of a PIK loan is usually feasible only after several years. PIK loans are often exposed to refinancing risk.

PIK toggle notes or bonds are a variant that allows the borrower to pay interest (partly or in full) in each period, or to accrue a part or the whole interest payment due. Usually if the borrower pays only part or nothing at all, the overall interest rate is increased following certain rules (often between 25 and 100 basis points). PIK toggle notes usually stipulate specific cash flow triggers that would effectively trigger interest payments during a certain period. PIK toggle notes often come with light covenants.

36.4.4.3 Subordinated Debt with Profit Participation A **subordinated debt with profit participation** scheme provides a risk balance between debt and equity to mezzanine lenders, offering a level of downside protection and also a way to participate in the upside potential. The following example presents a possible case of a profit participation scheme.

Suppose a £10 million loan is arranged as part of a profit participation model. The loan starts in January 2016, and principal will be repaid at maturity, which is in eight years. The profit participation scheme (PPS) is 4% of earnings before interest and taxes (EBIT), and annual floors and caps of £160,000 and £250,000, respectively, are set up. The profit participation amounts to be received by the lender during each year, assuming yearly sales and EBIT, are shown in Exhibit 36.16.

Exhibit 36.17 shows the respective calculations. For example, in 2016, the PPS provides a payment of £128,000 (i.e., 4% of £3,200,000). This is lower than the floor payment of £160,000; therefore, the floor would be binding in 2016. The same phenomenon happens in 2017 and 2018. However, in 2019, and reflecting the continuing rise in sales and EBIT of the previous years, the PPS is equal to £208,000, which is greater than the floor of £160,000 but lower than the cap of £250,000. Therefore, this is the payment for that year. The PPS payments for 2020 and 2021 are also between the cap and the floor, and thus these are the payments to be received by the lender (£240,000 in 2020 and £232,000 in 2021, as sales stabilized during

EXHIBIT 36.16 Sales and EBIT—Example of a Combined Profit Participation Model

Financials/Year	2016	2017	2018	2019	2020	2021	2022	2023
Sales	£10,000,000	£11,500,000	£13,000,000	£14,500,000	£16,000,000	£16,100,000	£18,100,000	£20,000,000
EBIT	£ 3,200,000	£ 3,500,000	£ 3,800,000	£ 5,200,000	£ 6,000,000	£ 5,800,000	£ 6,800,000	£ 7,300,000

Source: Authors' calculations.

EXHIBIT 36.17 Example of a (Combined) Profit Participation Model

Financials/Year	2016	2017	2018	2019	2020	2021	2022	2023
Sales	£10,000,000	£11,500,000	£13,000,000	£14,500,000	£16,000,000	£16,100,000	£18,100,000	£20,000,000
EBIT	£ 3,200,000	£ 3,500,000	£ 3,800,000	£ 5,200,000	£ 6,000,000	£ 5,800,000	£ 6,800,000	£ 7,300,000
PPS	£ 128,000	£ 140,000	£ 152,000	£ 208,000	£ 240,000	£ 232,000	£ 272,000	£ 292,000
Cap	£ 250,000	£ 250,000	£ 250,000	£ 250,000	£ 250,000	£ 250,000	£ 250,000	£ 250,000
Floor	£ 160,000	£ 160,000	£ 160,000	£ 160,000	£ 160,000	£ 160,000	£ 160,000	£ 160,000

$10 million PP loan starting in January 2016
PPS: 4% of EBIT
Tenor: 8 Years
Principal repayment at maturity

Floor: £160,000 annually
Cap: £250,000 annually

Source: Authors' calculations.

those two years). Finally, as sales resume growth in 2022 and 2023, the PPS for those two years (£272,000 and £292,000, respectively) exceed the cap payment of £250,000. Thus, PPS payments for 2022 and 2023 are capped at £250,000 in each year. Note that in Exhibit 36.17, the highlighted cells correspond to the actual PPS applicable to each year.

36.4.4.4 Subordinated Debt with Warrants None of the products presented so far may be attractive to a mezzanine lender when a firm has exhausted its capacity to assume further senior and subordinated debt and is experiencing limited predictability of its future cash flows. In such cases, to justify assuming the risk that comes with the investment, the lender will endeavor to obtain an equity exposure to the firm. A warrant is a security that fits perfectly in this situation, as it allows the investor to benefit from the potential upside in the equitable value of the firm.

Warrants are similar to equity options, but they differ in that they (1) are generally issued by unlisted firms and are thus regarded as OTC securities, (2) are dilutive when issued by the firm itself, (3) tend to have much longer maturities (often years) than equity options (which usually have maturities measured in months), and (4) are not standardized securities.

Warrants are often attached to bonds or preferred stocks as a sweetener, as they allow investors to potentially receive an extra return (coming from the warrant) above the coupons received from the bond (or the fixed dividend received in the case of a preferred stock). In turn, warrants allow borrowers to make considerably lower interest payments compared to subordinated loans of similar risk with no warrants attached. Warrants are different from other mezzanine securities because they eventually grant investors actual equity in the firm.

Take the following example: Firm ABC issues bonds with a $100 face value. The bonds also have warrants attached, providing each bondholder the right to purchase 10 shares of firm ABC stock at $15 per share over the next five years.

Equity warrants can be call and put warrants or a combination of both.

Most warrants are issued by private companies and do not trade in formal exchanges (they trade over the counter). This raises questions when using the Black-Scholes option pricing formula to value these securities, which is nonetheless the most widely accepted model used to value warrants.

36.4.4.5 Convertible Loans Convertible loans were introduced while discussing the convertible arbitrage hedge fund strategy. Here we discuss convertible bonds in the context of mezzanine financing. Traditionally, the convertible bond market has been regarded as a source of funds for companies with non-investment-grade credit ratings. However, the market has expanded over the past decade to include companies with investment-grade ratings. Convertible bonds are attractive to investors seeking to obtain equity exposure while enjoying downside protection.

The convertible bond market encompasses four types of debt instruments:

1. Publicly issued convertible bonds
2. Small-cap and micro-cap convertible bonds
3. Private investments, normally with specific conditions attached
4. Privately placed equity-linked products and restricted stock situations

A convertible bond is similar to a bond with a warrant attached. However, in the case of convertibles, the option is embedded in the fixed-income product. The exercise price (the conversion price) of the warrant embedded in a convertible is paid by handing over the underlying bond; therefore, the exercise price changes through time, complicating the valuation of these instruments. Also, the fact that most convertibles are callable creates another valuation problem.

Convertibles offer two distinct advantages to issuers: lower interest costs and less restrictive covenants compared to non-convertible bonds. Investors in convertibles are willing to accept lower yields and less restrictive covenants in exchange for receiving a more senior security that provides comparative income stability and the possibility of enjoying an increase in the value of the bond if the underlying value of the shares rises.

36.4.5 Mezzanine Financing and Project Finance

This final section on mezzanine financing briefly discusses the uses of mezzanine debt and financing in project finance. **Project** is capital often used for specific purposes, such as real estate projects and infrastructure projects, either on a private basis or in a public-private partnership. Public-private partnerships (PPPs) are discussed later in the section.

The majority of the funding in project finance (around 75%) comes from debt products. Therefore, the high cost of using standard fixed-coupon, long-term bonds (10 to 15 years) would render many of these projects financially unfeasible. In these cases, lenders will be willing to fund projects only if they can provide senior, non-recourse loans that are secured by all the assets of the project.

Lenders will also be able to assume control of a project if needed, as they will have a first lien on all assets and contracts. The cost of the funding is also usually reduced by improving the creditworthiness of the project by enhancing the collateral. On top of this, a performance bond will often be issued by an insurance company to guarantee that the contractors finish the project on time. The government can also guarantee a certain minimum price for the product or service to be offered (e.g., a minimum toll fare to be collected on a project regarding the building of a highway). Finally, there is also the use of staged financing, whereby lenders finance the development of a project up to a certain point. The financing for further stages will become available only when certain parts of the project are completed.

Public-private partnerships represent cooperation between governmental and business actors who work jointly toward a mutual target, assuming investment risks and sharing revenues and costs based on a predefined distribution. PPPs have become important participants in the project finance landscape and often use mezzanine finance products. The public authority (public party) specifies its requirements in terms of outputs, which set out the public services that the facility (public infrastructure) is intended to provide but do not specify how these are to be provided. It is then left to the private sector to design, finance, build, and operate the facility to meet these long-term output specifications. The project company (private party) receives payments (from the public-sector party or from the general public as users of the facility) over the life of the PPP contract on a pre-agreed-upon basis, which are intended to repay the financing costs and give a return to investors. The facility remains in public-sector ownership, or it is reverted to public-sector ownership at the end of the PPP contract. PPP projects are used to build, maintain, and operate an extensive infrastructure deployed in the public interest using the financial resources of the private sector, where not only the financial resources of private enterprises but also their know-how and organizational and innovative potentials are used (Svedik and Tetrevova 2012, 114).

PPP projects have the following characteristics:[10]

- Projects consist of a partnership between two or more participants from the public and private sectors.
- The partnership is based on a long-term relationship among participants, which have a shared responsibility in the results of their actions.
- Each actor is a principal and can negotiate on its own behalf.
- Participants must be prepared to devote material and non-material resources in the partnership.
- Projects are usually long-term and thus require long-term financing, often from alternative financial vehicles (belonging to debt, equity, and mezzanine financing sources).
- Projects entail a significant financial scale, and the invested risk capital is usually borne solely by the private partner.

PPP project financing is usually arranged by creating a special purpose vehicle and designing a risk-sharing, cash flow–based lending structure in which limits are set to liabilities and off-balance-sheet financing. Projects are financed through a mixture of equity and debt financing, with debt financing in the neighborhood of 70% or more.

36.5 CONCLUSION

Since its inception two decades ago, the market for insurance-linked securities (both non-life and life) has experienced a healthy rate of growth, facilitating the transfer of insurance risk. This risk can be transferred to investors for the following reasons: (1) to transfer catastrophe risk (a large-scale catastrophe could wipe out many insurance companies at the same time); (2) to avoid overexposure to a few reinsurers, thus lowering credit risk; and (3) in the case of life insurance, to turn it into a tradable security (since, as explained, life settlements allow policyholders to sell their life insurance policies).

In the case of cat bonds, some expect that climate change may unleash increasingly costly natural catastrophes and hence the need to cover this risk. In the case of life insurance products, according to a report by the Basel Committee (2013), an underestimation of life expectancy by one year causes an estimated increase of 3% to 4% in the value of the liabilities that defined benefit funds have assumed. This highlights the importance of being able to hedge longevity risk. Longevity derivatives are an example of a product that can be used to hedge this risk.

Pension funds and other institutions are particularly exposed to the risk of policyholders living longer than expected. The low interest rate environment that has prevailed since the financial crisis of 2008–9 has exacerbated the financial impact of longevity risk on pension plans and other providers of retirement income. This is because lower discount rates increase the present value of the cost of longevity risk events. It is expected that longevity-related products will continue to gain acceptance by pension funds and other institutions to hedge longevity risk. This will lead to the design of new products and to improvements in existing ones. In this new environment, the modeling of mortality risk will be of utmost importance.

The last part of the chapter was devoted to mezzanine finance products, all of which have a debt and an equity component to them. In the aftermath of the financial crisis, many financial institutions entered into mezzanine lending, and interest in mezzanine products is expected to remain strong in the future.

NOTES

1. Cat bond returns were measured using the Swiss Re Cat Bond Total Return Index, world equity returns reflect the total returns of MSCI's World Equities Index, and global bond returns were measured using JPM's Aggregate Global Bond Index Total Returns.
2. Cat bond returns were measured using the Swiss Re Cat Bond Total Return Index, world equity returns reflect the total returns of MSCI's World Equities Index, global bond returns were measured using JPM's Aggregate Global Bond Index Total Returns, U.S. high yield is based on returns to Barclays' U.S. Corporate High Yield, and commodity returns reflect the returns to the S&P GSCI.
3. See Bodoff and Gan (2009).
4. See Gaches (2012). Luciano and Regis (2014) report that the largest-ever longevity swap transaction has been for a notional amount of £5 billion (the buyer of the contract was Aviva and the seller was Swiss Re and SCOR).
5. See OSFI (2014).
6. This section and the following are based mainly on Krutov (2010).
7. See Bussmann (2011) and Robinson et al. (2011).
8. Silbernagel and Vaitkunas (2012).
9. This section is based mainly on Nijs (2012).
10. See Bult-Spiering and Dewulf (2006) and Svedik and Tetrevova (2012).

REFERENCES

Basel Committee. 2013. "Longevity Risk Transfer Markets: Market Structure, Growth Drivers and Impediments, and Potential Risks." Bank for International Settlements, December.

Biffis, E., and D. Blake. 2010. "Securitizing and Tranching Longevity Exposures." *Insurance: Mathematics and Economics* 46 (1): 186–97.

Bodoff, N., and Y. Gan. 2009. "An Analysis of the Market Price of Cat Bonds." Casualty Actuarial Society E-Forum, Spring.

Braun, A., N. Gatzert, and H. Schmeiser. 2011. "Performance and Risks of Open-End Life Settlement Funds." Working Papers on Risk Management and Insurance 73, March.

Bult-Spiering, M., and G. Dewulf. 2006. *Strategic Issues in Public-Private Partnerships*. Oxford, UK: Blackwell Publishing.

Bussmann, D. 2011. "Financing Infrastructure in Asia: Mezzanine Capital." 5th Annual Infrastructure Investment World Asia, October 19.

Center for Entrepreneurship, SMEs and Local Development. 2013. "Alternative Financing Instruments for SMEs and Entrepreneurs: The Case of Mezzanine Finance." CFE/SME(2012)9/FINAL, Organisation for Economic Co-operation and Development, February 8.

Coburn, A., and P. Nakada. 2012. "Understanding Longevity Risk: Insights from Structural Cause-of-Improvement Model." In "Longevity Risk Management for Institutional Investors." Special issue, *Institutional Investors Journal* (Fall): 92–99.

Credit Suisse. 2006. "Mezzanine Finance: A Hybrid Instrument with a Future." *Economic Briefing* 42.

Cummins, J. D. 2008. "Cat Bonds and Other Risk-Linked Securities: State of the Market and Recent Developments." *Risk Management and Insurance Review* 11 (1): 23–47.

Edesess, M. 2014. "Catastrophe Bonds: An Important New Financial Instrument." EDHEC–Risk Institute, July.

Gaches, A. 2012. "Index-Based Longevity Swaps: The Next Big Thing?" In "Longevity Risk Management for Institutional Investors." Special issue, *Institutional Investors Journal* (Fall): 64–68.

Guy Carpenter & Company, LLC. 2015. "Catastrophe Bond Update: Fourth Quarter 2014."

International Monetary Fund (IMF). 2012. "The Financial Impact of Longevity Risk." Chap. 4 in *Global Financial Stability Report: The Quest for Lasting Stability*. International Monetary Fund, Washington, DC, April.

Jaeger, L., S. Müller, and S. Scherling. 2010. "Insurance-Linked Securities: What Drives Their Returns?" *Journal of Alternative Investments* 13 (2): 9–34.

Januário, A., and N. Naik. 2014. "Testing for Adverse Selection in Life Settlements: The Secondary Market for Life Insurance Policies." http://bvlzl.de/Newsletter/2014_12/Markets_USA_01.pdf.

Krutov, A. 2010. *Investing in Insurance Risk*. London: Risk Books.

Kusche, P. 2013. "The New Alternative Asset: Insurance-Linked Securities." *Journal of Structured Finance* 19 (2): 52–58.

Luciano, E., and L. Regis. 2014. "Risk-Return Appraisal of Longevity Swaps." In "Pension and Longevity Risk Transfer for Institutional Investors." Special issue, *Institutional Investors Journal* (Fall): 99–108.

NAIC. 2012. "A Comprehensive Overview of the Insurance-Linked Securities Market." National Association of Insurance Commissioners, Capital Markets Special Reports.

Nijs, L. 2012. *Mezzanine Financing*. Croydon, UK: Wiley Finance.

OSFI. 2014. "Longevity Insurance and Longevity Swaps." Office of the Superintendent of Financial Institutions Canada. www.osfi-bsif.gc.ca.

Robinson, A., I. Fert, and M. Brod. 2011. "Mezzanine Finance: Overview." Practical Law Company.

Silbernagel, C., and D. Vaitkunas. 2012. "Mezzanine Finance." Mercer, Bond Capital Mezzanine Inc.

Sterge, A., and B. Van der Stichele. 2015. "Understanding Cat Bonds." AQR Capital Management.

Svedik, J., and L. Tetrevova. 2012. "Financing and Mezzanine Capital in the Context of PPP Projects in the Czech Republic." In *Recent Researches in Business and Economics: Proceedings of the 4th WSEAS World Multiconference on Applied Economics, Business and Development (AEBD '12)*, edited by Zeljko Panian, 113–17. Athens: WSEAS Press.

Swiss Re. 2012. "The Fundamentals of Insurance-Linked Securities." Swiss Re Corporation. www.swissre.com/.

Weistroffer, D. 2010. "Insurance-Linked Securities: A Niche Market Expanding." Deutsche Bank Research.

APPENDIX A

Alternative Presentations of Mean-Variance Optimization

There are other ways to present and implement mean-variance optimization. The first approach involves the choice of an optimal portfolio in which the expected return is maximized subject to a set of constraints. More specifically,

$$\begin{aligned}
&\max_{w_i,\ldots,w_N} \sum_{i=1}^{N} w_i E[R_i] \\
&\text{Subject to} \\
&\text{Var}\left[\sum_{i=1}^{N} w_i E[R_i]\right] \leq v \\
&\sum_{i=1}^{N} w_i = 1 \\
&a_i \leq w_i \leq b_i
\end{aligned} \tag{A.1}$$

In this optimization, the expected rate of return of the portfolio, $\sum_{i=1}^{N} w_i E[R_i]$, is maximized by selecting weights, w_i for $i = 1, \ldots, N$, subject to a number of constraints. First, the variance of the portfolio must be less than v. Second, the weights must add up to one, and then each weight should satisfy a set of lower and upper limits.

Alternatively, the optimization can be expressed in terms of minimizing the variance of the portfolio subject to a set of constraints. That is,

$$\begin{aligned}
&\min_{w_i,\ldots,w_N} \text{Var}\left[\sum_{i=1}^{N} w_i E[R_i]\right] \\
&\text{Subject to} \\
&\sum_{i=1}^{N} w_i E[R_i] \geq \mu \\
&\sum_{i=1}^{N} w_i = 1 \\
&a_i \leq w_i \leq b_i
\end{aligned} \tag{A.2}$$

Both statements will lead to the same set of optimal portfolios and efficient frontier.

Index

Bold page numbers indicate keyword definitions or explanations.