FINANCIAL DERIVATIVES

Pricing and Risk Management

Robert W. Kolb, James A. Overdahl, Editors

KOLB SERIES IN FINANCE

Essential Perspectives

FINANCIAL DERIVATIVES

The *Robert W. Kolb Series in Finance* provides a comprehensive view of the field of finance in all of its variety and complexity. The series is projected to include approximately 65 volumes covering all major topics and specializations in finance, ranging from investments, to corporate finance, to financial institutions. Each volume in the *Kolb Series in Finance* consists of new articles especially written for the volume.

Each *Kolb Series* volume is edited by a specialist in a particular area of finance, who develops the volume outline and commissions chapters by the world's experts in that particular field of finance. Each volume includes an editor's introduction and approximately 30 articles to fully describe the current state of financial research and practice in a particular area of finance.

The chapters in each volume are intended for practicing finance professionals, graduate students, and advanced undergraduate students. The goal of each volume is to encapsulate the current state of knowledge in a particular area of finance so that the reader can quickly achieve a mastery of that special area of finance.

FINANCIAL DERIVATIVES

Pricing and Risk Management

Robert W. Kolb
James A. Overdahl

The Robert W. Kolb Series in Finance

WILEY

John Wiley & Sons, Inc.

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

Library of Congress Cataloging-in-Publication Data:

Kolb, Robert W., 1949–
Financial derivatives : pricing and risk management / Robert W. Kolb, James A. Overdahl.
p. cm. – (The Robert W. Kolb series in finance)
Includes bibliographical references and indexes.
ISBN 978-0-470-49910-8 (cloth)
1. Derivative securities. 2. Financial engineering. I. Overdahl, James A. II. Title.
HG6024.A3K648 2010
332.64′57–dc22 2009017152

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Introduction

In a time in which the finance industry is under attack and our entire financial system is under remarkable stress, financial derivatives are at the center of the storm. For the public at large, financial derivatives have long been the most mysterious and least understood of all financial instruments. While some financial derivatives are fairly simple, others are admittedly quite complicated and require considerable mathematical and statistical knowledge to understand fully.

With vast changes for our financial system in prospect, there has never been a time in which those engaged in setting public policy and the concerned general public have a greater need for a general understanding of financial derivatives. As the reader of this book will learn, financial derivatives are instruments of remarkable power and very justifiable uses. However, as this text also freely acknowledges and explains, the very power of these financial derivatives makes them subject to accident in the hands of the incautious and also makes them effective tools for mischief in the hands of the unscrupulous.

To contribute to an improved public understanding of these markets, *Financial Derivatives* explores the contemporary world of financial derivatives, starting with a presumption of only a general knowledge of undergraduate finance. These chapters have been written by many leading figures in academics, industry, and government for the benefit of advanced undergraduates, graduate students, practicing finance professionals, and the general public. As such, the chapters in this book provide a comprehensive understanding of financial derivatives. *Financial Derivatives* is comprised of 37 chapters organized into six parts:

Part One, "Overview of Financial Derivatives," provides an introduction to and an overview of the types of financial derivatives, the markets in which they trade, and the way that traders use derivatives, and it also offers a broader perspective addressing the question of the social function of derivatives markets. Against that background, Part Two, "Types of Financial Derivatives," explores the variety of derivatives, starting with the agricultural and metallurgical derivatives that were historically the first to be developed. This part also discusses financial derivatives based on stock indexes, foreign currencies, energy, and interest rate instruments. It continues by giving an overview of the variety of exotic options and a type of exotic options known as an event derivative. Two chapters focus on credit default swaps and structured credit products that have allegedly played a central role in the recent crisis in financial markets. Executive compensation is always controversial, it seems, and has generated particular outrage in the current crisis, so this part discusses executive stock options and concludes with an overview of some of the emerging financial derivatives that are likely to become prominent in the future.

After having introduced the markets and types of derivatives in Parts One and Two, Part Three turns to an examination of "The Structure of Derivatives Markets and Institutions." Chapter 17 analyzes the development and current state of derivatives markets, and subsequent chapters take on issues such as a survey of the participants in the market and the way in which transactions are fulfilled. Fulfillment is a critical part of the market, because this issue concerns the honoring and completion of contracts, without which no viable market can persist. Closely related to this is the issue of counterparty credit risk—the risk that one party to the derivatives contract might default on contractual obligations. This part also surveys the regulation of derivatives markets, along with the principles of accounting as they pertain to derivatives. The part concludes with a brief account of some of the most famous derivatives disasters of recent decades.

Part Four, "The Pricing of Derivatives: Essential Concepts," introduces the fundamentals of determining the price of derivatives. The part begins by introducing the principle of no-arbitrage pricing. The first condition of a well-performing market from the point of view of pricing is that prices in the market are such that arbitrage is impossible—where arbitrage can be defined as the securing of a riskless profit without investment. With this background, the discussion turns to the pricing of particular instruments, such as forward and futures contracts. Next the part introduces the famous Black-Scholes option pricing model and then considers the various ways in which this seminal model has been extended and enhanced to apply to other derivatives. The part concludes with an analysis of the pricing of swap contracts.

Part Five, "Advanced Pricing Techniques," extends the pricing analysis initiated in Part Four. The chapters in this part are more technical, beginning with showing how Monte Carlo methods can be applied to price derivatives. The discussion of Monte Carlo techniques is immediately followed by a consideration of finite difference models, models that can be applied with great benefit when analytical models are not available. Much of the pricing of derivatives turn on the path that the underlying good is presumed to follow. When this path is described statistically, the description is known as a stochastic process, an understanding of which is necessary to more sophisticated analysis. Finally, this part explores how option prices respond to changes in their various input values.

Part Six, "Using Financial Derivatives," concludes the book. By this time, the reader will be well aware that financial derivatives are very valuable for managing risks and for providing information about the future prices of underlying goods. Financial derivatives can also be used as tools of quite sophisticated speculation. This part begins with an exploration of option strategies used in speculation and shows how the same strategies can also be used to reduce risk. Next comes a discussion of how hedge funds use financial derivatives and, more exactly, how hedge funds use the techniques of financial engineering. Financial derivatives are powerful tools for managing interest rate risk, as this part also explores. Chapter 36 examines real options, options based on physical assets or opportunities that firms possess. The book concludes with a discussion of how firms can use financial derivatives to manage their own risks.

Acknowledgments

The editors would like to acknowledge the contribution of the many people who have made this volume possible. Our first debt is to the many scholars who shared their knowledge by writing the chapters that comprise this text. We would like to also thank George Lobell, editor at John Wiley & Sons, Inc., for his vision of the series in which this volume appears and his encouragement of the series in general and this text in particular. Also at John Wiley, we would like to offer our thanks to the editorial team of Pamela Van Giessen, William Falloon, and Laura Walsh for their continuing support of and commitment to this project.

PART I

Overview of Financial Derivatives

Part One consists of four introductory chapters intended to open the world of financial derivatives to the reader. In Chapter 1, "Derivative Instruments: Forwards, Futures, Options, Swaps, and Structured Products," Gary D. Koppenhaver takes a generalist approach to forwards, futures, swaps, and options. He approaches these instruments from the point of view of their suitability to address a single problem: managing financial risk. Through this approach, he shows that these instruments obey common principles and are closely related from a conceptual point of view. Koppenhaver strives to emphasize the connections among these different types of derivatives in order to demystify derivatives in general.

One of the largest differences among derivatives turns on the manner in which they are traded—on exchanges or in the more informal and less structured over-the-counter market? Sharon Brown-Hruska contrasts these two models for trading derivatives in Chapter 2, "The Derivatives Marketplace: Exchanges and the Over-the-Counter Market." In light of the financial crisis, many legislators are pressing to reduce or eliminate the over-the-counter market, which is actually much larger than the market for exchange-traded derivatives. However, many believe that trading derivatives on exchanges make them more transparent, easier to regulate, and less likely to lead to derivatives disasters.

From the point of view of derivatives, we might think of *speculation* as trading derivatives in a manner that increases the investor's risk in order to pursue profit. *Hedging* by contrast is trading derivatives in order to reduce a preexisting risk. In Chapter 3, "Speculation and Hedging," Gregory Kuserk shows how hedging and speculation differ but also explains how one might think of hedging and speculating as two sides of the same coin, with the relationship between the two activities being much closer than is generally recognized.

The editors of this volume believe that Chapter 4 by Christopher L. Culp, "The Social Function of Financial Derivatives," is one of the most important in the entire volume. As discussed in the introduction to this book, there is a recurring impulse to eliminate derivatives markets through legislative action. Culp shows how derivatives markets serve society in a variety of ways, some of which are quite obvious and others of which are more sophisticated.

CHAPTER 1

Derivative Instruments

Forwards, Futures, Options, Swaps, and Structured Products

G. D. KOPPENHAVER
Professor and Chair, Department of Finance, Insurance and Law,
Illinois State University

INTRODUCTION

The evolution of ideas in finance usually is driven by circumstances in financial markets. In the early 1980s, at the inception of cash-settled financial futures contracts, the term *derivatives* was most often associated with financial rocket science. Esoteric derivative contracts, especially on financial instruments, faced a public relations probem on Main Street. By the mid-1990s, the term *derivatives* carried a negative connotation that conservative firms avoided. High-profile derivative market losses by nonfinancial firms, such as Metallgesellschaft AG, Procter & Gamble Co., and Orange County, California, caused boards of directors to look askance at derivatives positions.[1] In the early 2000s, however, derivatives and their use are a real part of a discussion of business tactics. While it is still the case that derivatives contracts are a powerful tool that could damage profitability if used incorrectly, the discussion today does not focus on why derivative contracts are used but how and which derivative contracts to use.

The goal of this chapter is to take a generalist approach to closely related instruments designed to deal with a single problem: managing financial risk.[2] In the chapter, forwards, futures, swaps, and options are not treated as unique instruments that require specialized expertise. Rather the connection between each class of derivative contracts is emphasized to demystify derivatives in general. As off–balance sheet items, each is an unfunded contingent obligation of contract counterparties. Later in the chapter, the discussion returns full circle to consider the creation of funded obligations with derivative contracts, called structured products. Structured products are financial instruments that combine cash assets and/or derivative contracts to offer a risk/reward profile that is not otherwise available or is already offered but at a relatively high cost. The repackaging of off–balance sheet credit derivatives into an on–balance sheet claim is shown through a structured investment vehicle example.

Uncertainty is a hallmark of today's global financial marketplace. Unexpected movements in exchange rates, commodity prices, and interest rates affect earnings and the ability to repay claims on assets. Great cost efficiency, state-of-the-art production techniques, and superior management are not enough to ensure firm profitability over the long run in an uncertain environment. Risk management is based on the idea that financial price and quantity risks are an ever-increasing challenge to decision making. In responding to uncertainty, decision makers can act to avoid, mitigate, transfer, or retain a commercial risk. Because entities are in business to bear some commercial risk to reap the expected rewards, the mitigation or transfer of unwanted risk and the retention of acceptable risk is usually the outcome of decision making. Examples of risk mitigation activities include forecasting uncertain events and making decisions that affect on–balance sheet transactions to manage risk. The transfer of unwanted risk with derivative contracts, however, is a nonintrusive, inexpensive alternative, which helps explain the popularity of derivatives contracting.

Consider Exhibits 1.1 through 1.4 as part of the historical record of volatility in financial markets. Exhibit 1.1 illustrates the monthly percentage change in the Japanese yen/U.S. dollar exchange rate following the breakdown of the Bretton Woods Agreement in the early 1970s. The subsequent exchange rate volatility helped create a successful Japanese yen futures contract in Chicago. In Exhibit 1.2, the monthly percentage change in a measure of the spot market in petroleum is illustrated. While significant spikes in price occur around embargos or conflict in the Middle East, price volatility has not lessened over time for this important input to world economies. U.S. interest rates are also a source of uncertainty. The change in Federal Reserve operating procedures in the late 1970s temporarily increased volatility, but significant uncertainty in Treasury yields has remained over time.

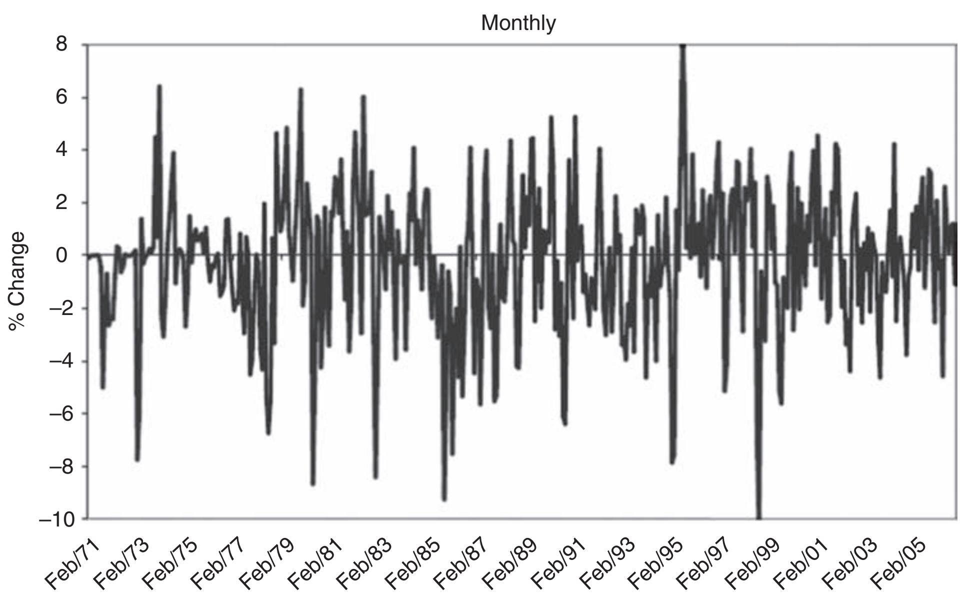


Exhibit 1.1 Percent Change in Yen/U.S. $ Exchange Rate

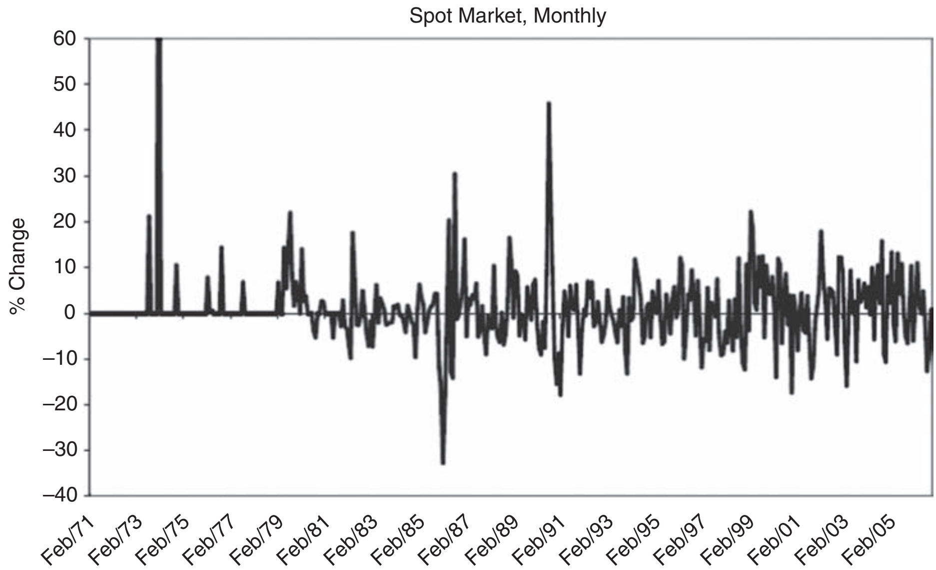


Exhibit 1.2 Percent Change in West Texas Oil Prices

Exhibit 1.4 illustrates the past history of default risk premiums. Most recently, a sharp spike in default risk premiums occurred at the end of the stock market technology bubble in the early 2000s. Across all graphs, it should be clear that uncertainty in economically important markets is not decreasing over time and that the effectiveness of forecasting changes in prices, rates, or spreads as a method to mitigate the uncertainty is not likely to be high.

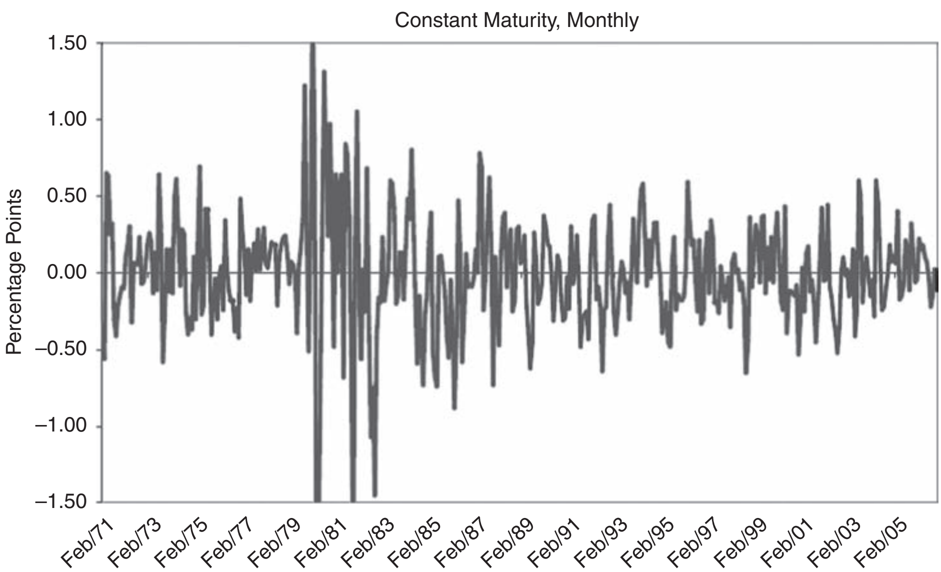


Exhibit 1.3 Change in 5-Year Treasury Yield

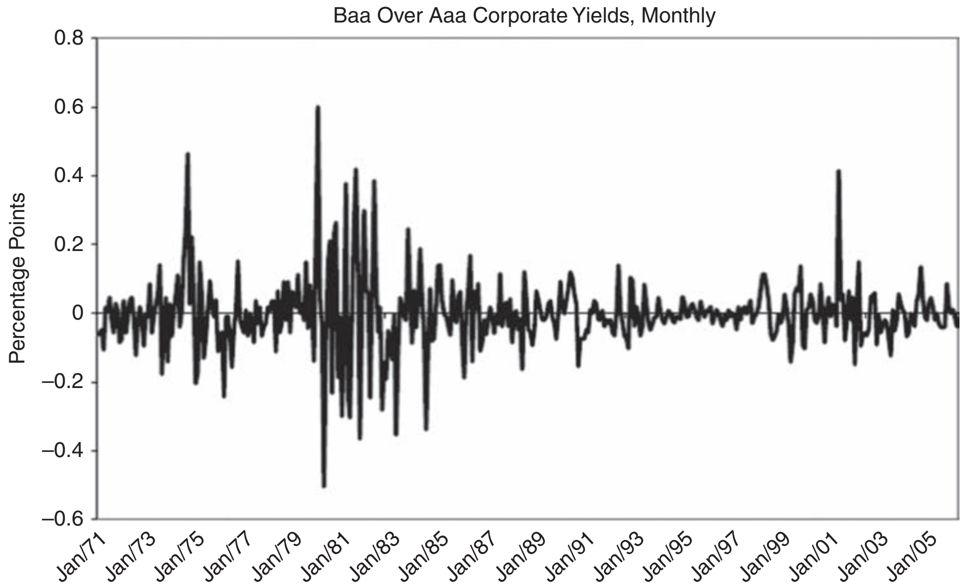


Exhibit 1.4 Credit Risk Premium Change

A GENERALIST'S APPROACH TO DERIVATIVE CONTRACTS

What are derivative contracts? A derivative contract is a delayed delivery agreement whose value depends on or is derived from the value of another, underlying transaction. The underlying transaction may be from a market for immediate delivery (spot or cash market) or from another derivative market. A key point of the definition is that delivery of the underlying is delayed until sometime in the future. Economic conditions will not remain static over time; changing economic conditions can make the delayed delivery contract more or less valuable to the initial contract counterparties. Because the contract obligations do not become real until a future date, derivative contract positions are unfunded today, are carried off the balance sheet, and the financial requirements for initiating a derivative contract are just sufficient for a future performance guarantee of counterparty obligations.

Before beginning a discussion of contract types, it is helpful to depict the profiles of the commercial risks being managed with derivative contracts. The first step in any risk management plan is to accurately assess the exposure facing the decision maker. Consider Exhibit 1.5, which plots the expected change in the value of a firm, ΔV, as a function of the unexpected change in a financial price, ΔP. The price could be for a firm output or for a firm input. The dashed line indicates that as the price increases ($\Delta P > 0$) unexpectedly, the value of the firm falls. The specific relationship is consistent with many conditions, such as an unexpected rise in input cost, a loss of significant market share as output prices unexpectedly rise, or even a rise in the price of a fixed income asset due to an unexpected decline in yields. The key is simply that the unexpected price rise causes the expected value of the enterprise to fall.

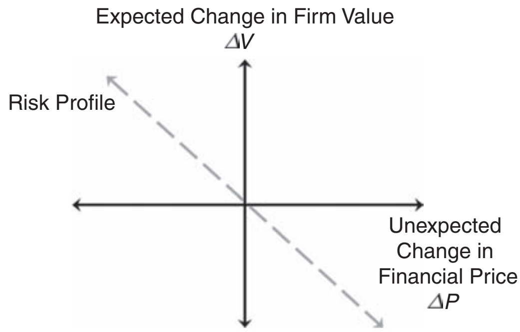


Exhibit 1.5 A Commercial Risk Profile

It is also instructive to ask whether there are alternatives to derivative contracts in managing commercial risks. Significant, low-frequency commercial risks are transferred through insurance contracts, for example. While virtually any risk can be insured, negotiation costs and hefty premiums may prevent insurance from being a cost effective mechanism for risk transfer. On–balance sheet transactions such as the restructuring of asset and/or liability accounts to correct an unwanted exposure are another alternative to derivative contracts. Customer resistance to restructuring may affect profitability as, say, a squeeze on net interest income results when a bank offers discounts on loans or premium deposit rates to accomplish the restructuring. Finally, firms can exercise their ability to set rates and prices to transfer risk to customers and stakeholders. Such exercise of market power as an alternative to derivative contracting depends on the degree of competition in output and input markets. Firms facing different competitive pressures may have different preferences for derivatives relative to other risk transfer methods.

Forward Contracts

The most straightforward type of derivative contract is a contract that transfers ownership obligations on the spot but delivery obligations at some future date, called a forward contract. One party agrees to purchase the underlying instrument in the future from a second party at a price negotiated and set today. Forward contracts are settled once—at contract maturity—at the forward price agreed on initially. Industry practice is that no money changes hands between the buyer and seller when the contract is first negotiated. That is, the initial value of a forward contract is zero. As the price of the deliverable instrument changes in the underlying spot market, the value of a forward contract initiated in the past can change.

To illustrate the value change in a forward contract, consider Exhibit 1.6. All other things equal and for every unexpected dollar increase in the financial price, ΔP, an agreement to purchase (long forward) the underlying instrument at the lower forward price increases expected firm value, ΔV. Alternatively, Exhibit 1.6 shows that an agreement to sell (short forward) the underlying instrument at the lower forward price decreases expected firm value. The forward contract long (short) benefits from the contract if the underlying instrument price rises (falls) before the contract matures. The exhibit also shows that both buying and selling

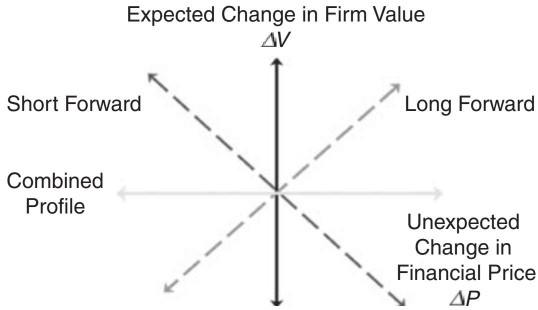


Exhibit 1.6 Risk Profiles for Forward Contracts

exactly the same forward contract create a combined position that makes the firm value insensitive to unexpected changes in the underlying price (the horizontal axis). Comparing Exhibits 1.5 and 1.6, the commercial risk profile in Exhibit 1.5 is the same as the risk profile for a short forward contract. To hedge away the risk or make the firm insensitive to unexpected changes in the underlying price, the firm should enter into a long forward contract (Exhibit 1.6).

A feature of a forward contract is that the credit or default risk implicit in delayed delivery performance is two-sided. The default risk is real because most forward contracts are settled by physical delivery. Recall the forward contract buyer can either make a gain or take a loss depending on the forward price set initially and the price of the underlying at contract maturity. If the underlying instrument price rises (falls), the contract buyer gains (loses) on the forward contract. Because the value of the contract is settled only at contract maturity and no payments are made at origination or during the term of the contract, a forward contract buyer is exposed to the credit risk that the seller will default on forward contract delivery obligations when the underlying asset can be sold for more in the spot market. Likewise, a forward contract seller is exposed to the credit risk that the buyer will default on forward contract payment obligations when the underlying asset can be purchased for less in the spot market.

Consider a forward rate agreement as an example of a forward contract on interest rates. A forward rate agreement is an agreement to pay a fixed interest rate on a pre-determined, notional principal amount and receive a floating rate cash flow on the same notional principal amount at contract maturity. Note that only the interest cash flows are intended to change hands at contract maturity. If the floating rate return is higher than the fixed rate cost agreed to at contact initiation, the forward rate long gains the difference in cash. If the floating rate return is lower than the fixed rate cost agreed to at contact initiation, the forward rate short gains the difference in cash. The forward rate long gains if interest rates rise or fixed income prices fall over the life of the contract. A map of the forward rate agreement cash flows is illustrated in Exhibit 1.7, where $\bar{R}$ is the fixed rate set at contract origination, time 0, and $\tilde{R}$ is the actual rate realized at time t, the maturity of the contract.

Suppose three months in the future Ford Motor Acceptance Corporation (FMAC) plans to borrow \$100 million for three months at the U.S. dollar

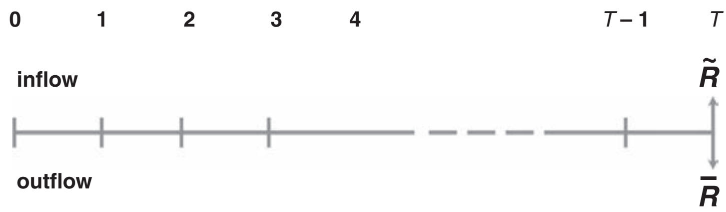


Exhibit 1.7 Forward Rate Agreement Cash Flows

London Interbank Offered Rate (LIBOR). FMAC is exposed to the risk that borrowing costs (rates) will rise unexpectedly over the three months. If so, the \$100 million amount borrowed will raise less cash for use by FMAC. The commercial risk facing FMAC is illustrated in Exhibit 1.8, which is similar to Exhibit 1.6 except that the unexpected change affecting firm value is a change in interest rates instead of prices. FMAC decides to manage the interest rate risk by making a long forward rate agreement with an investment bank as seller. What should the fixed rate be on FMAC's forward rate agreement if LIBOR 3-month is 4.9507 percent and LIBOR 6-month is 5.1097 percent? A "fair" forward rate would be one that does not favor either the buyer or the seller of the forward rate agreement nor create an opportunity for interest rate arbitrage. The three-month rate three months in the future from the LIBOR yield curve is 5.2036 percent annualized $(=\{[1 + (.051097 \times 182/360)/(1 + (.049507 \times 91/360))] - 1\} \times (360/91))$. That is, LIBOR rates must rise from 4.9507 percent to 5.2036 percent for investors to be indifferent between a sequence of two three-month LIBOR investments and one six-month LIBOR investment yielding 5.1097 percent. Let the forward rate agreement specify 5.2036 percent as the fixed rate. If the three-month LIBOR rate in three months is greater than 5.2036 percent, FMAC receives a net payment on the forward rate agreement from the investment bank to offset the greater liability issuance costs associated with the increase in LIBOR rates.

Futures Contracts

Because of unique, negotiated features in forward derivative contracts, the contracts can be difficult to reverse or terminate early once created. Futures contracts

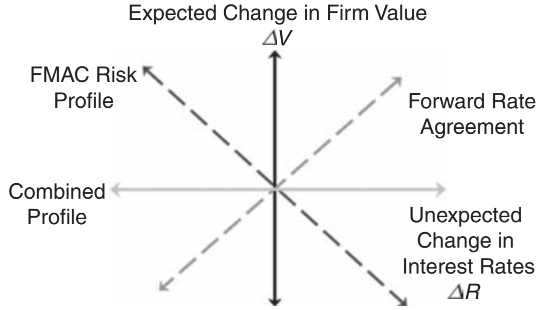


Exhibit 1.8 Risk Management with a Forward Rate Agreement

are delayed delivery contracts that permit the frequent transfer of ownership and delivery obligations until some future date. Once created, a futures contract can be settled in one of two ways: delivery of the underlying at contract maturity (less than 2 percent of contracts) or, more commonly, liquidation of a prior position by an offsetting transaction. A centralized marketplace allows investors to trade contracts with each other, providing immediacy in the execution of transactions, and the regulation of trade practices. Because contract sellers are obligated to perform only in the future, there are no restrictions on short selling as in the cash markets for equity, for example.

Futures contract trading has many unique institutional features compared to trading the underlying spot instruments or forward contracts. As a financial safeguard, earnest money deposits are made by both futures contract buyer and futures contract seller to ensure performance of future obligations. As unfunded positions, the funds required to trade futures contracts are a performance bond, not an amount borrowed from the broker to make up the entire contract value. The earnest monies are usually less than 10 percent of futures contract value. Futures contract positions are marked to market at least daily, which means position value changes are conveyed from losers to gainers through performance bond adjustments when prices change. The constant transfer of funds and removal of financial claims between traders effectively reduces the trader's performance period.

Default risk is further reduced by a clearinghouse for all transactions; the clearinghouse is organized to provide a third-party guarantee of financial performance for all cleared trades. In effect, the clearinghouse becomes the seller to every buyer and buyer to every seller, standing behind or guaranteeing each transaction. Most important, futures contracts are standardized, not customized, to apply to certain underlying commercial risks. The contracts specify the quantity or par value of the underlying instrument and whether the contract is settled by cash or physical delivery at contract maturity, among other things. Contract standardization and clearinghouse operation make taking a futures position that offsets or eliminates a previous position a low-cost alternative to holding an open contract until maturity. In total, these institutional features decrease the cost of contract origination and early termination.

In comparing futures to forward contracts, note that the risk profiles are the same (see Exhibit 1.9). That is, long and short futures positions have the same effect on firm value change as long and short forward positions, respectively. The success of futures markets as a risk-transfer device is largely due to their cost advantage compared to forward contracts. Importantly, futures contracts involve less credit risk exposure for traders. Marking futures contracts to market prices at least daily removes debt from the marketplace in risk transfer, which makes default less likely.

Futures contracts can also be considered an extension of forward contracts on the same underlying instrument. Suppose two counterparties enter into a sequence of one-day forward contracts. The forward contract is negotiated first on day 0 and settled on day 1. For simplicity, assume the forward contract buyer pays the difference between the forward price and the day 1 spot market price if the forward price is larger than the spot price. If the forward price is less than the spot price at day 1, the seller pays the difference to the buyer. A new forward contract is then written on day 1 reflecting the day 1 spot price, again maturing in a single day. If the

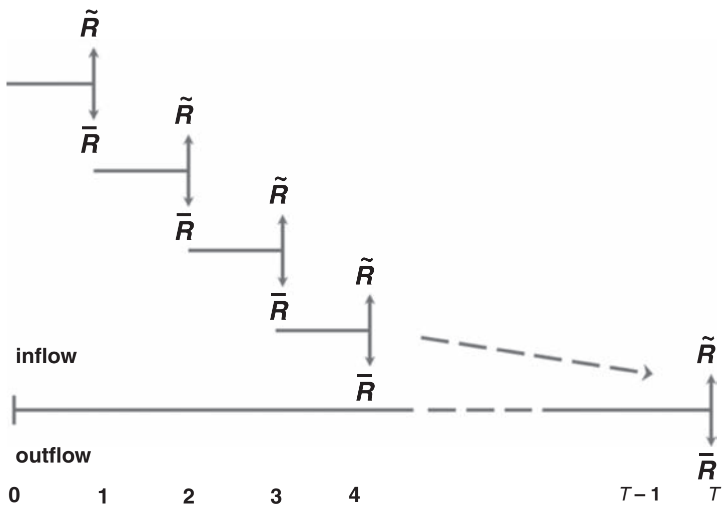


Exhibit 1.9 A Futures Contract as a Portfolio of One-Day Forward Contracts

sequence of forward contracting (forwards are negotiated, mature, are renegotiated and repriced daily) is repeated until some future time t (futures contract maturity), the result is a futures contract on the same underlying instrument. The entire "portfolio" of one-day forward contracts captures the institutional safeguard of marking all futures positions to market at least daily, removing the financial claims that can be built up over the life of a single forward contract. Futures contracts can be viewed as just a sequence of forward contracts.

Swap Contracts

A swap contract is a contract in which two counterparties agree to make periodic payments that differ in a fundamental way from each other until some future date. The terms of a swap contract, besides the maturity and notional value of the contract, can include the currencies to be exchanged (foreign currency swap), the rate of interest applicable to each counterparty (interest rate swap), and the timetable by which payments are made. Swap contracts are an over-the-counter, negotiated derivative contract like a forward contract rather than an exchange-traded instrument like a futures contract. The swap contract counterparties must be classified as Eligible Contract Participants, as defined by the Commodity Exchange Act.[3] Although foreign currency swaps predate interest rate swaps, interest rate swaps are economically most important today.

Consider the uses for an interest rate swap. Suppose the swap contract specifies the exchange of floating rate cash flows for fixed rate cash flows. That is, a counterparty agrees to pay a fixed cash flow (based on a fixed rate) to another counterparty and in return receive a variable cash flow (based on a floating rate). The exchange of cash flows occurs periodically, say every six months; the cash flows are netted against each other so that whichever counterparty's cash flow is larger, that counterparty pays the difference to the other. A swap of fixed rate for floating rate cash flows reduces the fixed rate payer's exposure to unexpected rate increases, which is an important commercial risk for the holder of existing fixed-income securities or firms that anticipate the issuance of debt in the future. If rates rise during the contract's life, the fixed rate payer will receive cash flows that offset the loss of value in existing securities or the increase in debt issuance cost. Similarly, the swap of floating rate for fixed rate cash flows reduces the floating-rate payer's exposure to unexpected rate decreases.

Suppose in 1999, Maytag issues a three-year, $100 million par floating rate note with semiannual interest payments at 80 basis points over the London Interbank Offered Rate (LIBOR), which is at 3.2 percent per annum. Maytag's interest expense floats with LIBOR but at the current rate, Maytag will pay $2.0 million as interest to the debt holders every six months, in March and September. Maytag is exposed to the commercial risk of unexpected increases in LIBOR. To transfer the risk, Maytag enters into a swap agreement with a U.S. commercial bank to pay a fixed 5 percent per annum on the $100 million until maturity. In return, the bank agrees to pay Maytag a variable amount based on LIBOR plus 80 basis points. Only the cash flow differential is exchanged in the agreement. Exhibit 1.10 is a table of the swap cash flows as LIBOR rises. Maytag makes increasing cash payments to debt holders as the rate floats higher; completely offsetting the interest expense are equivalent cash inflows from the U.S. commercial bank. Maytag still pays a fixed rate cash flow to the U.S. commercial bank of $2.5 million every six months. Note that Maytag has credit risk exposure from the bank only when the value of the interest rate swap is positive (last column of Exhibit 1.10).

Because the performance period for swaps (six months in the above example) is generally less than for forward contracts (only at contract maturity) but greater than for futures contracts (at least daily), the default risk characteristics of swap contracts are intermediate between forward and futures contracts. At each time when cash flows are exchanged between counterparties, the value of the interest rate swap is effectively marked-to-market and reset to zero. Between interest rate reset dates (again, a six month period in the above example), the obligation to pay a known fixed rate and receive an unknown variable rate depends on the movement

		---------Millions of Dollars---------		
		FLOATING	*FIXED*	Net
Date	Rate	Cash Flow	Cash Flow	Cash Flow
Mar. 1, 1999	4.2%			
Sept. 1, 1999	4.8%	+2.10	–2.50	–0.40
Mar. 1, 2000	5.3%	+2.40	–2.50	–0.10
Sept. 1, 2000	5.5%	+2.65	–2.50	+0.15
Mar. 1, 2001	5.6%	+2.75	–2.50	+0.25
Sept. 1, 2001	5.9%	+2.80	–2.50	+0.30
Mar. 1, 2002	6.4%	+2.95	–2.50	+0.45

Exhibit 1.10 Swap Contract Cash Flows for Maytag

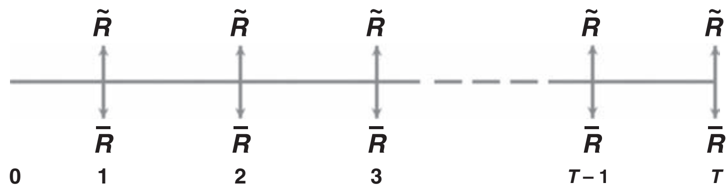


Exhibit 1.11 A Swap Contract as a Portfolio of Different Maturity Forward Contracts

in rates. The swap contract builds a financial claim by one counterparty against the other between reset dates, creating either a positive or negative swap value, as interest rates change. This is similar to a forward contract once it is negotiated but not yet settled. To carry the connection to forward contracts further, a swap contract can be viewed as a portfolio of different maturity forward contracts, each maturing at a different swap interest reset date (see Exhibit 1.11). The Maytag interest rate swap in our example then consists of six forward rate agreements, the first maturing and settled in six months, the last maturing and settled in three years.

Viewing a swap as a portfolio of different maturity forward contracts has two important implications. First, a swap contract has a risk profile similar to a forward contract or a futures contract on the same underlying instrument. That is, the fixed rate payer in the typical interest rate swap has a position that protects against an unexpected rise in interest rates, like the forward rate agreement in Exhibit 1.8 or a short futures position in Eurodollar time deposits. If rates rise and prices fall unexpectedly, the additional cash inflow in the swap contract or a forward rate agreement or a Eurodollar futures contract offsets the increased interest expense. The floating rate payer has the opposite risk profile. The second implication is that a portfolio of forward contracts with different maturities (a swap contract) can be valued on the assumption that today's forward interest rates are realized. The value of the swap is just the sum of the values of the forward contract elements of the swap. Knowing how to price forward contracts and value forward contract positions is all that is needed to price swap contracts and value swap positions between rate reset dates.

Option Contracts

Option contracts fall into one of two basic categories: calls or puts. In a call (put) option contract the contract buyer has the right but not the obligation to purchase (sell) a fixed quantity from (to) the seller at a fixed price before a certain date. Every option contract has both a buyer and a seller. The contract buyer has a right but not an obligation to initiate an exchange; the seller is obligated to perform, however, should the buyer exercise the contract rights. The fixed price in an option contract is the exercise or strike price—the price at which the contract buyer either purchases from the contract seller (call option) or sells to the contract seller (put option). The contract maturity date is also called the contract expiration date. Finally, the option buyer makes a nonrefundable payment to the option seller, called the option premium, to obtain the rights of the option contract. The purpose of an option pricing model, such as the Black-Scholes model or the binomial model, is to estimate a "fair" option contract premium.

In general, a call option buyer (seller) expects the price of the underlying security to increase (decrease or stay steady) above the option exercise price. If not, the call option seller keeps the nonrefundable payment, the call option premium. A put option buyer (seller) expects the price of the underlying security to decrease (increase or stay steady) below the option exercise price. If so, the put option buyer can exercise the right to sell the underlying instrument to the put option seller at the relatively high exercise price. If an option contract is held to expiration, the option may expire worthless, be exercised by the contract buyer, or be sold for the difference between the contract exercise price and the market price of the underlying.

Consider the call option risk profile in Exhibit 1.12. The buyer of an option contract, call or put option, is called the option long; the option seller is called the option short. In Exhibit 1.12, if the unexpected change in the underlying instrument's price, ΔP, at option expiration is negative (or prices fall), the long call position is worthless and the call option buyer forfeits the call premium. At the same time, the short call option position is profitable by the amount of the premium. The horizontal, dashed lines to the left of the vertical axis illustrate the returns. If the unexpected change in the underlying instrument's price, ΔP, at option expiration is positive (or prices rise), the long call position increases the value of the option buyer, ΔV. Before the option buyer can break even, however, the price must rise sufficiently to cover the nonrefundable option premium paid to the option short. At the same time, the short call position keeps part of the premium paid by the call long until prices rise sufficiently. The sloping, dashed lines to the right of the vertical axis illustrate the returns. Exhibit 1.12 shows that the risk profile of a long call position is similar to a long forward or long futures contract position. The risk profile of a short call option position is similar to a short forward or futures contract position but only if underlying prices rise.

In Exhibit 1.13, the risk profiles for a put option are illustrated. Because a put is an option to sell the underlying instrument, unexpected price increases, $\Delta P > 0$, result in a constant return equal to the put option premium for the option short or loss of the same for the option long. That is, the option to sell at a relatively low price is worthless if prices rise unexpectedly. The horizontal, dashed lines to the right of the vertical axis illustrate the returns. To the left of the vertical axis,

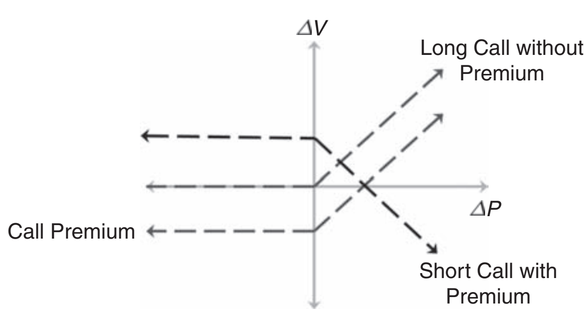


Exhibit 1.12 Risk Profiles for Call Options

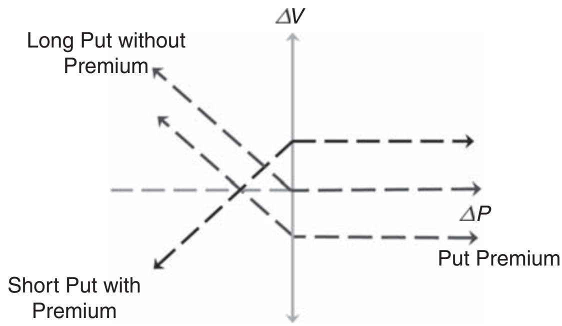


Exhibit 1.13 Risk Profiles for Put Options

the put long gains and the put short loses dollar for dollar as the underlying price falls unexpectedly. Exhibit 1.13 shows that the risk profile of a long put position is similar to a short forward or short futures contract position. Note that although the put option buyer is long the option, the position is effectively short the underlying instrument. The risk profile of a short put option position is similar to a long forward or long contract position but only if underlying prices fall.

Next consider the risk profile of positions that combine a commercial risk with a call option. Suppose we return to Maytag's problem in 1999. Recall Maytag issued floating rate debt tied to LIBOR and so is exposed to the risk that interest rates rise unexpectedly. Maytag's commercial risk profile is the same as the risk profile illustrated in Exhibit 1.14. That is, as interest rates increase the additional interest expense of funding at LIBOR plus 80 basis points decreases the value of the firm. To manage the risk, Maytag purchases a call option on LIBOR interest rates from an investment bank. If rates fall in the future, Maytag will not exercise its right to issue debt at the exercise rate and will forgo the option premium. If rates rise in the future, Maytag will gain from the call option by issuing debt at below-market rates. The call option gain offsets the higher interest expense to ensure that the most that can be lost is the option premium. The combined risk profile in Exhibit 1.14 shows how Maytag's commercial risk is transformed after the purchase of a call option on LIBOR. Most important, note that the combined profile in this exhibit is the same

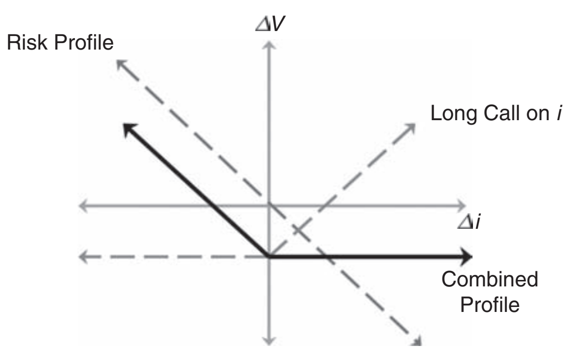


Exhibit 1.14 Risk Management with Call Option Contracts

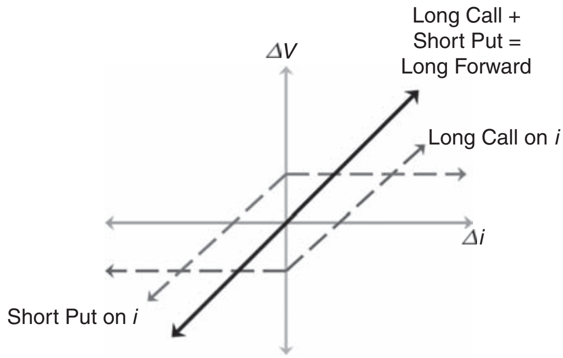


Exhibit 1.15 Relationship between Options and Forward Contracts

risk profile for a long put option position (Exhibit 1.13), suggesting a relationship between call options, put options, and the underlying instrument. The relationship is termed put-call parity.

Let us generalize the relationship of option contracts to forward contracts. For an increase in financial value, a call option contract has a risk profile like a forward contract on the same underlying instrument. For a decrease in financial value, however, a call option has a risk profile like a Treasury bill with the same maturity as the option, yielding a constant return if held to maturity. A call option risk profile can then be replicated by adjusting the composition of a portfolio of a forward (or futures or swap) contract on the underlying instrument and a riskless security. An analogous relationship holds for a put option contract and a short forward (or futures or swap) contract for a decrease in financial value. A put option risk profile can also be replicated by adjusting the composition of a portfolio of a forward (or futures or swap) contract on the underlying instrument and a riskless security. To carry the relationship a step further, suppose a portfolio combines a long call contract and a short put contract on the same underlying instrument with the same exercise prices and expiration date. The resulting risk profile is depicted in Exhibit 1.15. As shown, the risk profile of a long forward contract position is replicated by the portfolio of a long call contract and a short put contract on the same underlying instrument with the same exercise prices and expiration date.

STRUCTURED PRODUCTS AND AN APPLICATION TO DERIVATIVE CONTRACTS

Given a basic understanding of derivative contracts and how derivative contracts are related, consider structured products with an application to derivative securities. Structured products are financial instruments that combine cash assets and/or derivatives to provide a risk/reward profile not otherwise available or only available at high cost in the cash market.

Structured products are often the securities that result from the securitization process and have been successfully created with portfolios of mortgage, automobile, and boat loans as well as credit derivatives. Securitization is the process by which mortgages issued by a financial institution are converted into mortgage-backed securities sold to investors or a firm's accounts receivable portfolio is converted into asset-backed notes, for example. Investors in securitized loans face default and prepayment risk, among other risks, passed through by original loan issuer. Usually investors are offered multiple classes—called tranches—of notes that differ in their priority of payment rights to best match the investor's risk-bearing preferences. Depending on the securitization structure, investors may choose between super-senior notes (first to be repaid, last to default), mezzanine notes (second to be repaid, second to default), and equity or capital notes (last to be repaid, first to default). Each tranche earns a different return based on the structured risk.

Of course, the correspondence of corporate securities and option contracts is familiar to students of derivatives.[4] A zero-coupon bond issued by a firm without other debt can be viewed by the bondholder as a riskless bond plus a written put option (default premium) on the value of the firm's assets. The put option is owned by shareholders and the firm can be "put" to the bondholders in exchange for a release of indebtedness. With multiple tranches in a structured product, junior debt claims are paid only after senior debt claims. The super-senior tranche, for example, can still be viewed as a riskless bond plus a written put option. The mezzanine and capital note claims can be viewed as bullish spreads (put options both purchased and sold) on the value of the issuing firm's assets. Note that the correspondence between option contracts and corporate securities applies to all default-risky debt instruments.

To apply the ideas behind structured products to derivative contracts, consider an example using a recently developed credit derivative building block: a credit default swap. That is, these derivative contracts have the credit quality (or default risk) characteristics of a single company (called a single-name credit default swap) or a portfolio of debt instruments from different companies (called a portfolio default swap) as the underlying instrument. A change in the reference entity's default risk characteristics causes the value of the credit derivative to change. The International Swaps and Derivatives Association (ISDA), a trade group for the swaps industry, sponsors master agreements that define the situations under which a payoff is triggered by a reference entity default.[5] Because debt instruments underlie a credit derivative, borrowers (firms, governments) and lenders (financial intermediaries, investors) might use a credit derivative to manage the risk of an unexpected change in a reference entity's default risk. The risk of issuing or holding any fixed income security includes interest rate risk as well as credit risk. In essence, a credit derivative contract represents a low-cost way to separate credit risk from interest rate risk. There are and have been well-accepted "credit derivatives" like variable-rate loan commitments, standby letters of credit, revolving loan facilities, and floating rate loans (floaters), which are all affected by changes in the credit quality of the funding beneficiary. As always, the key to the successful acceptance of credit default swaps is their low transaction costs.

Specifically, a credit default swap is not so much a "swap" of cash flows as an "option" on credit quality. In a credit default swap, the contract buyer pays a periodic premium to the contract seller and the contract seller protects the buyer from unexpected changes in the credit risk of a reference asset. The buyer and seller agree to a definition of a credit event; if the credit event occurs, the protection seller

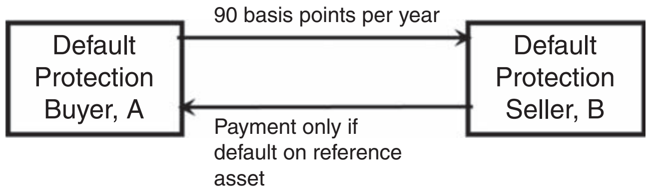


Exhibit 1.16 Cash Flows for a Credit Default Swap

pays the contract buyer a predetermined amount and the contract is terminated. If there is no credit event over the life of the derivative contract, the protection seller keeps the premium cash flow paid by the protection buyer. From a relatively simple arrangement with a single firm's debt as the reference entity, complicated agreements can be constructed, such as a total return swap and a portfolio default swap. A general illustration of the payments in a credit default swap is shown in Exhibit 1.16. The analogy between a credit default swap and an insurance policy against adverse changes in default risk should be clear.

To return to the topic of this section, a credit default swap is not a structured product per se. A credit default swap, however, can be the raw material behind a structured product.[6] As with most derivative contracts, the value of a credit default swap is initially set to zero: Neither the protection buyer nor the protection seller is at an advantage over the other. That is, the buyer agrees to pay periodic premiums to the seller that reflects the current default risk of the reference entity. The credit default swap seller has an unfunded liability contingent on the occurrence of a credit event. Suppose that the protection seller in effect securitizes the unfunded liability by issuing debt instruments called credit-linked notes to investors in the cash market. Securitization means the protection seller repackages the payment rights to the protection buyer's premium and sells those rights to investors. Investors in the notes purchase cash market debt instruments that are linked to the default risk of the reference entity underlying the credit default swap. Investors receive the premium cash flow from the credit default swap, which typically is augmented with income from collateral purchased with the proceeds of the securitization, for the risk of a credit event. The credit-linked notes tied to the credit default swap and purchased by investors in the cash market are one example of a structured product.

Why securitize a derivative contract based solely on default risk? Certain institutional investors may be prohibited by policy or regulation from taking direct positions in credit derivative contracts yet have permission to invest in notes with traditional coupon payments, par values, and credit ratings. Credit-linked notes may have a transaction cost advantage over negotiated credit default swaps yet the same diversifying exposure to default risk. Finally, credit-linked notes are valuable as an instrument to hedge other credit derivatives positions.

In essence, the issuance of a structured note offsets the exposure of selling a credit default swap, which creates a funded rather than unfunded exposure. The investor in a credit-linked note pays for cash flows associated with selling the protection of a credit default swap. If no credit event occurs in the reference entity, the investor (indirectly, the protection seller) keeps the premium cash flow. If a credit event occurs, however, the investor's claim on the cash flow and collateral investment income is junior to at least the protection buyer in the credit default swap. Structured products can repackage the cash flows of credit default swaps, broadening the market for credit derivatives in general. The credit-linked notes can be viewed generally as corresponding to options (default-risky securities) on credit default swaps (generating income passed through to note holders). Finally, the securitization process using the safest, super-senior notes, for example, could repeat itself, meaning the ultimate investors that provide the funding for the credit protection are several steps removed from the original seller of the credit default swap.

CONCLUSION

This chapter takes a generalist approach to instruments designed to deal with a single problem: managing financial risk. Forwards, futures, swaps, and options are not unique instruments that require specialized expertise. The connection between each class of derivative contracts helps demystify derivatives in general. Viewing forwards, futures, swaps, and options as interrelated tools for risk transfer may help further the application of integrated risk management. Interest rate, exchange rate, credit, prepayment, and price risks are often interrelated. Financial markets and commercial risks are not becoming simpler, more local, or less important than in the past. The goal of the chapter has been to provide a framework to make the connections clear. Finally, the creation of funded obligations with derivative contracts, one example of a structured product, also is an important application of the connection between derivative contracts. The repackaging of off–balance sheet credit derivatives into an on–balance sheet claim, illustrated with a structured investment vehicle that issues credit-linked notes, shows that derivative contracts are a flexible, essential part of the risk management landscape.

ENDNOTES

1. For an excellent discussion of the possible lessons learned by successful and unsuccessful uses of derivative contracts, see Marthinsen (2005).
2. The idea that derivative contracts are fundamentally interrelated was first detailed effectively in Smithson (1998). The approach here draws significantly from that treatment.
3. An eligible contract participant is an entity, such as a financial institution or investment trust, classified by the Commodity Exchange Act on the basis of assets or regulated status to engage in transactions not generally available to noneligible contract participants. Noneligible contract participants include retail customers and individual investors, for example.
4. The discussions in Chapter 7 of Cox and Rubinstein (1985) and Chapter 15 of MacDonald (2006) are excellent treatments of the options view of security design.
5. Credit default swap contracts sponsored by the ISDA specify rights and obligations of the counterparties and definitions of a reference entity credit event. Reference entity default is commonly triggered by bankruptcy filings, a failure to pay on notes, or a debt moratorium, restructuring, or repudiation. See Chapter 24 of Bomfim (2005) for detailed discussion of IDSA contracts.
6. Precisely speaking, the example of a structured product discussed here is called a synthetic collateralized debt obligation. The term *synthetic* is used because a derivative contract—a credit default swap—is being securitized instead of a portfolio of cash market debt instruments. See Bomfin (2005) and Rosen (2007).

REFERENCES

Bomfim, A. 2005. *Understanding Credit Derivatives and Related Instruments*. Amsterdam: Elsevier Academic Press.

Cox, J., and M. Rubinstein. 1985. *Options Markets*. Englewood Cliffs, NJ: Prentice-Hall.

MacDonald, R. 2006. *Derivatives Markets*. Boston: Addison-Wesley.

Marthinsen, J. 2005. *Risk Takers: Uses and Abuses of Financial Derivatives*. Boston: Pearson Addison-Wesley.

Rosen, R. 2007. "The Role of Securitization in Mortgage Lending." Chicago Fed Letter, Federal Reserve Bank of Chicago 224 (November).

Smithson, C. 1998. *Managing Financial Risk: A Guide to Derivative Products, Financial Engineering, and Value Maximization*. New York: McGraw-Hill.

ABOUT THE AUTHOR

G. D. Koppenhaver is currently the chairperson of the Department of Finance, Insurance and Law at Illinois State University. Dr. Koppenhaver has held a joint appointment as a professor in the Finance Department and in the Economics Department at Iowa State University. He was employed by the Federal Reserve Bank of Chicago as a senior economist and has been a consultant to the Federal Reserve Bank of Dallas. Dr. Koppenhaver was also a faculty member at Southern Methodist University.

Dr. Koppenhaver's research interests include the optimal use of futures, options, and swap contracts to manage risk and the operations of money market mutual funds. His research has appeared in *Decision Sciences Journal of Innovative Education*, *Financial Education and Practice*, *Journal of Banking and Finance*, *Journal of Finance*, *Journal of Financial Research*, *Journal of Financial Services Research*, *Journal of Futures Markets*, *Managerial and Decision Economics*, and *Review of Futures Markets.*

CHAPTER 2

The Derivatives Marketplace

Exchanges and the Over-the-Counter Market

SHARON BROWN-HRUSKA
Vice President, National Economic Research Associates

INTRODUCTION

Derivative contracts can take the form of standardized contracts listed and traded on an exchange or bilateral agreements negotiated between counterparties in the over-the-counter (OTC) market. Various terms of OTC derivatives are privately negotiated. This gives rise to tremendous diversity in OTC contracts, but it also can expose counterparties to substantial liquidity risk and credit risk.[1] Exchanges, however, trade standardized contracts through a centralized structure that is organized to promote liquidity and to mutualize credit risk. To contend with sharp price movements and the potential for counterparty default, exchanges and OTC markets have developed contracting and risk-sharing mechanisms to accommodate the different character of intermediaries and products in their markets.

Expansion of products and innovation has led to significant growth in both exchange and OTC market activities. Futures and options exchanges experienced an unbroken stream of volume increases from 2000 through 2008, growing from over 1.31 billion contracts in 2000 to over 6.95 billion in 2008 (Futures Industry Association 2008). OTC derivatives volumes also grew significantly, with dealers reporting over 35,000 transactions in 2001 growing to over 288,000 in 2008 (International Swaps and Derivatives Association 2008a, p. 3). The notional value of all OTC derivatives exceeded $683 trillion in June 2008, with interest rate swaps the largest component, approaching $357 trillion in notional value (Bank for International Settlements 2008, p. A-103). Electronic OTC markets have also grown rapidly. For example, the Intercontinental Exchange, Inc. (ICE), a market that was formed in 2000, hosted 247 million contracts with a notional value of $5.5 trillion in 2008, representing a 974 percent change in notional value from 2004.[2]

In the quest to capture the growing demand for their products, exchange and OTC markets are competitors but are also highly interdependent. As OTC dealers have focused on designing products for specialized commercial risks and those emanating from the banking and financial services industry in which they operate, they simultaneously rely on the price transparency and liquid hedging vehicles available on exchanges. Exchanges have sought to co-opt some of the vast OTC product space through the development of trading platforms for products based on interest rate swap contracts (e.g., the Chicago Mercantile Exchange [CME] Group's Swapstream) and indices on credit default swaps (CDS) (e.g., the iTraxx index products listed on Deutsche Borse's Eurex). While previous efforts of exchanges to provide trade execution and clearing services for swaps often have struggled to gain traction, in the current financial markets crisis the exchange model is increasingly seen as a means to increase transparency and liquidity and to reduce counterparty credit risk in the OTC markets.

The rapid growth of derivative markets, along with the failures of large derivatives users, including American Insurance Group, Bear Stearns, and Lehman Brothers, has been accompanied by concerns about the risks these instruments pose not only to users but also to the financial system as a whole. While these risks can be found to varying degrees in all markets, their dynamic in OTC markets can be particularly acute due to the potential for illiquidity of more complex products and the related challenges of valuation and the assessment of credit risk. Policy makers have proposed that requiring OTC derivatives to trade on an exchange, or requiring dealers to utilize an exchange-style central counterparty (CCP) to clear them, would increase transparency and reduce risks. But difficult questions remain regarding whether the exchange model can be adapted to accommodate the OTC market and actually accomplish these benefits without sacrificing the innovation and customization that have expanded economic capacity to transfer and manage risk.

In products and in processing, exchanges and OTC dealers' rapport is one marked by synergistic as well as competitive forces. This competitive but complementary relationship between the exchanges and OTC dealers has at times led to controversy regarding the regulation of OTC markets and the organization of exchanges, levels of disclosure and market transparency, and how to best mitigate the risks inherent in financial contracting. This chapter discusses how the exchange and the OTC markets, driven by competitive forces, have evolved and in many ways converged in their offering of products for the transfer and assumption of risk. Exchanges have experienced significant consolidation, and both exchanges and OTC markets have moved toward more centralized electronic processes. As technology has facilitated the growth in products and the volumes of trades, this growth has heightened the demand for exchanges and OTC derivatives dealers to take account of and put in place processes that address default risk as well as systemic risks associated with the trade and use of derivatives. Techniques to mitigate these risks, including the employment of collateral and netting, are discussed, as are efforts to increase the transparency of OTC derivatives markets.

STANDARDIZATION VERSUS CUSTOMIZED PRODUCTS: DIFFERENCES IN STRUCTURE AND APPROACH

Exchanges and OTC derivatives dealers are aggressively innovative. Futures and options exchanges focus on designing products that appeal to many users, and their offering of standardized products has created efficiencies in information aggregation and the provision of liquidity. By providing a market that brings together buyers and sellers for its products, exchanges help users to minimize search and negotiation costs (Carlton 1984, p. 241). Futures exchanges are organized to facilitate liquidity by setting contract features that encourage competing intermediaries to provide immediacy, which attracts buyers and sellers whose demand for contracts to contend with the "price risks of volatility" leads them to trade frequently (Grossman and Miller 1988, p. 619).

The liquidity that characterizes exchange-traded derivatives stems primarily from their standardized features and a market structure that channels trading interest into homogeneous contracts. Futures contracts, for example, specify the quantity or the minimum size of the position, the minimum price variation or "tick" size, daily settlement and mark-to-market procedures, and the expiration date at which the underlying asset will be delivered or the contract expires and final payments are made. While some precision in hedging may be lost in the form of basis or correlation risk when the standard product does not precisely match the underlying risk to be hedged, transactions and search costs are reduced as fewer elements of the transaction need to be negotiated.

Exchanges are characterized by strong network economies in that liquidity attracts order flow, and this, in turn, leads to greater liquidity and still lower costs. The information inherent in trades and the competition it engenders increases the efficiency of prices and further lowers the cost of trading, giving rise to economies of scale and scope. Pirrong (2008) observes that exchanges achieve these economies in the execution function, when orders to buy and sell are matched and priced in the exchange auction process, and in the clearing function, when the price and contractual obligations are confirmed and deliveries and payments are processed. As exchanges have shifted from a model in which all transactions had to pass through intermediaries on a trading floor to a fully electronic market in which everything from order entry to final settlement is processed digitally, they have been able to expand their product line and realize economies of scope and scale in liquidity and processing (Domowitz 1995).

Technology has also changed the OTC markets, but its impact has been greatest in the application of computer modeling and data in the design and pricing of products and risks that had previously been unavailable. Dealers pioneered the fields of financial engineering and structured finance, designing derivatives and structured products that range from those that are quite standardized, such as plain vanilla interest rate swaps, to highly customized and complex transactions focused on transforming risk and return for investors and commercial users. For example, while credit risk has long been a necessary risk of banking and business, the parsing and packaging of that risk into CDS, which pay off in the event of default or other credit event, created an efficient means to price and transfer that risk.

The development and expansion of credit derivatives paralleled the explosion of interest rate swap products a decade earlier, enabling financial institutions and commercial entities to recognize and manage default and credit risk with greater precision. Yet even as CDS products provided a means to transfer credit risk, they also increased the ease with which entities could assume greater levels of credit risk in exchange for increased return. As CDS trading increased, various operational and liquidity challenges characteristic of more customized instruments in the young product market would confront the swaps markets.

Due in part to the customized nature of OTC transactions, banks or broker-dealers are often necessary conduits for these privately negotiated derivative contracts. These OTC intermediaries have structured products that can increase the precision of the contracts to hedge or speculate on particular risks, isolate risks that are bundled in a business activity, or construct a portfolio of risks. To address risks and frictions associated with bilateral contracting, OTC intermediaries have adopted practices to reduce risks on those products for which they structure and act as a counterparty. In 1985, OTC market participants organized a trade association, the International Swaps and Derivatives Association (ISDA), in order to develop documentation and legal opinions to ensure the enforceability of contracts. ISDA's efforts, which are widely supported and encouraged by members of the financial community, banking authorities, and regulators, focus on addressing contingencies through enforceable legal documentation and best practices embedded in the documentation. To mitigate the risks inherent in private contracting, OTC dealers have sought to address contingencies through malleable features in standard contractual agreements, common definitions, and legal documentation.

The foundation for OTC contracts is the ISDA Master Agreement (or ISDA Master), which lays out the interpretations, general obligations, and agreed-on provisions governing the conduct of the derivatives transactions entered into between signatory counterparties. The goal of the ISDA Master is to provide certainty that contracts negotiated and agreed to will be fulfilled or, in the event of default or termination, are properly and smoothly closed out in order to minimize costs and risks borne by the respective counterparties. The ISDA Master governs all transactions between the counterparties, with the economic terms of the individual contracts set forth in a trade confirmation. Other supporting documents have been developed by ISDA dealers, including collateral provisions and credit documents (referred to as credit support annexes) and protocols that apply to certain events. Together with the standard definitions of the products and their accepted terms, these documents have come to be known collectively as the ISDA documentation architecture (Harding 2002).

Despite the importance and breadth of the common infrastructure developed by ISDA, the increased volumes and nonstandard features of OTC contracts and the performance obligations created by them can entail operational challenges. Concerns regarding documentation disparities and delays in confirmations led dealers to form the Counterparty Risk Management Policy Group II (CRMPG II).[3] In its 2005 report, the CRMPG II identified documentation management and operational efficiency as areas in need of improvement in light of the dramatic increases in derivatives volumes. OTC dealers recognized that automation, including electronic trade affirmation and matching, along with straight-through processing, or the integration of trading, reporting, and control functions through electronic media, were even more necessary. Working cooperatively with ISDA and other trade associations, OTC dealers have sought to increase the use of electronic automation in the processing of OTC transactions and have worked to reduce confirmation backlogs and strengthen back-office operations.

To address operational demands of increased volumes in OTC derivatives, the Depositary Trust and Clearing Corporation (DTCC), the Society for Worldwide Interbank Financial Telecommunications (SWIFT), SwapsWire, Euronext.liffe, and ICE developed stand-alone offerings that provide dealers with the ability to generate and reconcile trade confirmations in OTC derivatives. The DTCC Deriv/Serv application uses an auto-matching approach in which each counterparty generates its own trade record from back-office systems and enters it into the DTCC Trade Information Warehouse, where it is matched via an automated process with the results and any discrepancies notified to both parties for resolution. By turning to electronic applications that accomplish auto-matching in the confirmation and affirmation process, OTC dealers have significantly reduced backlogs and their associated operational risks.

The need to address the concerns raised by the CRMPG II and the credit crisis that began to unfold in early 2007 was made more urgent by the bankruptcy of Lehman Brothers and the unraveling of Bear Stearns. The response of dealers highlights how the approaches employed by exchange and OTC markets to mitigate the risks inherent in derivatives transactions have converged, with an increased reliance on electronic processes to achieve necessary results. In addition, the different approaches put forth by the exchanges and the OTC dealers reflect the competitive forces at work in these markets, to which we turn next.

COMPETITION AND CONSOLIDATION: IMPETUS FOR CHANGE

In a statement accompanying its decision not to block the merger of the CME and the Chicago Board of Trade (CBOT), the U.S. Department of Justice (DOJ) noted: "The two principal impetuses for innovation have been, and will continue to be, the prospect of winning business from the over-the-counter market and the potential to offer products that the OTC community can use to hedge the risk associated with its activities" (DOJ 2007, p. 1). In many respects, the merger of the two dominant exchanges in the futures industry typifies the significant changes in the character of derivatives markets, driven by exchange demutualization and technological advances in information communication and processing. As exchanges have demutualized, decision making moved from the province of floor traders to shareholders, and exchange management has become better able to compete on various fronts, including the OTC market space.

While floor trading still exists for some agriculture contracts, energy, and complex options, over 82 percent of futures and options trading in the United States is now electronic,[4] and many regard the storied futures pit as fast becoming a relic of exchange history. Exchange trading has evolved from orders arriving via telephone and brokers routed to floor traders for execution in a futures pit, to an electronic market in which bids and offers are entered via an integrated screen-based front-end, exposed to trading interest, and executed in milliseconds against interested buyers and sellers. In addition to traditional hedgers, speculators, and the futures brokers who intermediate their orders, new users have entered, and the activity of market making has changed. The locals, floor brokers, and scalpers who used to maintain a presence on the trading floor signaling and shouting orders to buy and sell have been replaced by a cybertrading crowd, including algorithmic traders, trading arcades, and direct access brokerages that cater to these frequent traders.

The innovations that have taken place in the exchange and OTC derivative markets have been accelerated by the competitive drive to increase the volume of transactions processed and settled. For exchanges, the impetus for change from open outcry trading to electronic trading was fostered by the transformation of exchanges from member-owned, "mutualized" cooperatives to publicly traded, for-profit entities. Prior to their demutualization, exchanges were organized as nonprofit entities whose members held seats on the exchange that entitled them to exclusive trading privileges and access unavailable to nonmembers. In the mutualized structure, exchange governance emanated from exchange leadership and member committees, elected by seat owners, which governed everything from business conduct to disciplinary proceedings to determination of daily settlement prices. This combination of self-regulation and trading privileges led to conflicts of interest in decision making, necessitating supervision of exchange conduct and operations.[5]

The inertia of the mutualized structure of exchanges, in which members controlled and sought to preserve the trading floor, historically led to significant resistance to the introduction of technology into the trading process and to changes that would lead to disintermediation and loss of order flow by the floor. Those who would transact business off the trading floor, including economically similar products like OTC swaps, were viewed as unregulated competition that usurped order flow transacted on the exchange. Cases were filed to assert that rules prohibiting off-exchange futures transactions applied to OTC swaps, and exchanges passed rules to discourage members from trading away from the exchange or transferring open interest. Given this backdrop, the institutional structure of the floor-based exchange set the stage for an uneasy competitive struggle between the intermediaries who would offer OTC derivatives products and the exchange members who were determined that all such transactions should be executed through an exchange.

The regulatory framework enacted in 2000, the Commodity Futures Modernization Act (CFMA), brought forth a new era of competition by making it easier for traditional exchanges to innovate and introduce new products and providing legal certainty for OTC transactions and new types of markets. Integrated firms, whose business lines include brokerage, credit enhancement, and intermediation in exchange and OTC markets, backed ventures to compete with exchanges. For example, BrokerTec Futures Exchange in 2001, the U.S. Futures Exchange in 2004, and a recent venture, ELX Electronic Liquidity Exchange, have had equity backing from brokerage firms.[6] Competitive challenges have also been levied by foreign exchanges seeking access to national markets, brought closer by screen-based systems and high-speed processing and OTC intermediaries and financial service providers situated globally. In addition to electronic markets based on the traditional market design, new entrants have developed proprietary trading systems to accommodate more flexible and efficient processes demanded in their different lines of business. Since the CFMA enabled the creation of proprietary markets in those products that neither fell into traditional agriculture nor financial assets, early efforts to develop competing markets were in energy commodities.

Early OTC contract markets included Enron Online, Houston Street, and the ICE, all of which developed electronic markets for trading of bilateral energy contracts following the deregulation of wholesale gas and power markets. In the aftermath of Enron's bankruptcy, market participants migrated to the centralized platform offered by ICE, in part because of its open architecture and lower transactions costs associated with negotiating bilateral deals. ICE featured an à la carte model, offering market users the option to register their bilateral deals via ICE's E-Confirm, to execute a transaction via ICE's anonymous trading system, and/or to clear via its relationship with a third-party clearinghouse, LCH-Clearnet. Concerns regarding counterparty credit risk following the Enron collapse, along with the operational efficiency of the ICE model, helped propel the ICE to the forefront of electronic markets in the OTC space and energized traditional exchanges, such as the New York Mercantile Exchange, to offer expanded execution capabilities for energy swaps and clearing services to the OTC markets (Acworth 2005).

The specter of competition from electronic markets, many of which were domiciled globally, and the loss of volume that it could entail for floor traders also resulted in substantial consolidation among exchanges and their processes. In a highly controversial action that pitted exchange management against member clearing firms, many of whom held stakes in the Board of Trade Clearing Corporation (BOTCC), the CBOT and the CME signed an agreement in April 2003 that combined their clearing operations and would transfer all open interest from the BOTCC to the CME Clearing House. The move to drop its 78-year relationship with the BOTCC was precipitated by BOTCC's negotiations with Eurex, a foreign competitor that had formed a U.S. subsidiary and announced intentions to enter the market for interest rate futures in direct competition with the marquee CBOT products.

With the CME clearing over 85 percent of all futures transactions conducted in the United States, concerns regarding the concentration of open interest into a single clearinghouse were outweighed by economic and operational efficiencies generated by the arrangement (Durkin and Gogol 2003). The success of the CME-CBOT clearing arrangement, its ability to generate efficiencies while avoiding prior competitive instincts that historically had pitted the Chicago exchanges against each other, cemented an alliance between the two largest U.S. exchanges that eventually led to their proposed merger. Because the merger had combined two formidable competitors and raised concerns that they would have the power to foreclose competition, the antitrust authority of the U.S. DOJ launched an inquiry into the proposed merger.

While the DOJ eventually approved the merger based on the case made by the exchanges that their primary competition lay in the OTC markets as well as the lack of overlap in the products traded by the two exchanges, the review by the antitrust experts brought increased attention to the clearing structure and potential anticompetitive outcomes it engendered. In response to a request for comments by the Department of the Treasury on the U.S. regulatory structure, the DOJ filed a comment suggesting that the current structure of clearing futures contracts by U.S. exchanges potentially posed a barrier to competition in the market for futures products and services. In its introductory summary, the comment noted that

> *[T]he control exercised by futures exchanges over clearing services . . . has made it difficult for exchanges to enter and compete in the trading of financial futures contracts. If greater head-to-head competition for the exchange of futures contracts could develop, we would expect it to result in greater innovation in exchange systems, lower trading fees, reduced tick size, and tighter spreads, leading to increased trading volume. (DOJ, 2008, p. 1)*

The current structure of the U.S. futures exchanges is one in which the activities of execution, clearing, and settlement are performed by divisions of the exchange, with control and decision making emanating from a unified management and board. This structure, in which business functions are vertically integrated in the industrial organization context, had historically led to a management structure drawn from and focused on the interests of the floor trading community. Clearing firms, entities that provide brokerage and funding to futures customers, represented by the Futures Industry Association, have sought a more competitive environment for exchange products and has long advocated for a more decentralized or "horizontal" organization of exchanges.

Clearing firms have maintained that a horizontal, utility-style model that would treat financial futures contracts as fungible, like that which exists in the options markets, would result in greater competition in the provision of execution services. In the options markets, the Options Clearing Corporation (OCC) is owned and managed by the exchanges as a common pool resource such that any option transaction executed at any exchange can be cleared through the OCC. An alternative approach, dubbed "freedom to clear," envisioned competitive clearinghouses from which customers would be able to select to clear and settle their transactions. This model would require exchanges to give up transactions executed on their exchange for clearing by a separate clearing entity. Exchanges have resisted both approaches, recognizing that the loss of open interest would make it easier for order flow to migrate to another exchange, tipping the edge they enjoy due to network effects to a lower cost exchange.

Pirrong (2008) argues that the vertical integration of execution and clearing is a natural expression of an industrial organization that minimizes transactions costs through economies of scale in execution and economies of scope in clearing. In their analysis of the antitrust economics of commodities exchanges, Falvey and Kleit (2007) also acknowledge the network economies at work in the exchange model, but note that these same economies of scale, in addition to sunk costs that arise from developing facilities and obtaining regulatory approvals, can constitute barriers to entry. They document various cases in the post-CFMA environment where incumbent exchanges have temporarily lowered exchange fees and erected rules that made it more difficult for challengers to succeed. Falvey and Kleit show that these actions had the effect of frustrating competitive efforts, and eventually the exchanges prevailed in preserving their dominance in futures trading. Neither the DOJ nor the CFTC was inclined to revisit the competitive implications of the rapid consolidation of the industry. The CME Group continued to expand, with the acquisition of NYMEX, while ICE acquired the New York Board of Trade (NYBOT) and the Winnipeg Commodity Exchange, and moved to unify all of its futures operations into a vertically integrated structure with both execution and clearing.

The concerns expressed regarding the potential market power that results from the consolidation of exchanges and the vertically integrated structure have been largely drowned out by the financial market crisis. The exchange model has been advocated in response to the defaults and illiquidity in the credit markets, and regulators have moved aggressively to mandate CCP clearing for CDS transactions in particular. However, a review of the mechanisms employed by OTC market participants to mitigate credit and systemic risk reveals they are similar in important respects to that of their exchange cousins. The next section explores mechanisms used to manage risks in OTC and exchange markets more directly.

MOVING FROM BILATERAL TO MULTILATERAL RISK MANAGEMENT

As the volume in derivatives transactions outstanding has grown exponentially, so too have counterparty obligations to exchange payments and assets. OTC dealers and regulators concerned about the size of these obligations and the potential for systemic risk emanating from a failure in a settlement somewhere in the chain of transactions have pressed for more comprehensive mechanisms to manage risks associated with derivative transactions.

In the OTC market, operational responsibility for settlement and counterparty credit risk management resides with the counterparties themselves. In the exchange, the structure of centralized execution and clearing transfers the operational responsibility for settlement and counterparty credit risk to the exchange clearinghouse. The clearing process utilizing a CCP has long been a distinguishing feature of exchanges. By entering as the CCP to all transactions executed on an exchange, the clearinghouse combines the risk-reducing features of multilateral netting, margin, and mark to market with risk mutualization. As credit risk has become a more pertinent concern in the credit crisis, the exchange model for performing these functions is looked to by many as a solution, yet it is important to recognize that the practice of netting, the use of collateral, and the diversification of risk are well-established practices in both exchange and OTC markets.

Collateral in Exchange and OTC Markets

The growing use of collateral to mitigate credit risk perhaps most aptly typifies the convergence of OTC market agreements with exchange practices. When a contract is entered on an exchange, the counterparties are required to post collateral on a symmetric basis with the exchange clearinghouse. In the case of futures exchanges, clearing firms collect the collateral, known as initial margin, from counterparties to a transaction and deposit it at the clearinghouse. The clearinghouse sets the collateral requirement based on the contracts assumed by the traders, and determines the collateral required based on changes in the value of the position, depositing and withdrawing from the collateral account based on gains and losses on those positions, referred to as marking to market.

In the exchange clearing model, marking to market entails frequent periodic settlement, daily and even on an intraday basis during periods of heightened volatility. The margin requirements are generally set at a level that would accommodate one day's market move, using proprietary models that use historical or implied volatilities of market prices to ensure coverage. Such frequent marking to market lowers settlement risk and credit risk by ensuring that losses and gains are covered and do not accumulate. As a result, collateral requirements in exchange-traded products are low (typically less than 2 percent of notional value), enabling highly efficient use of capital.

While frequent mark to market is critical to the exchange counterparty risk management system, cash flows necessitated by frequent mark to market can entail significant cost and loss of precision in particular hedging strategies. If an international manufacturer wants to hedge expected monthly billings denominated in foreign currency over the coming year, it would want a swap with a monthly reset tailored to pay based on the amounts of those cash flows using the floating exchange rate on a monthly basis. Cash flow hedging, a common structure used by small and large businesses across all economic sectors, is less efficient and would require more intermediation to accomplish under more frequent marking to market. For various types of risks faced in commercial and business activity, OTC contracts can be more efficient and less costly hedging vehicles than exchange contracts.

The OTC market structure enables counterparties to negotiate the frequency with which positions are marked to market and variation collateral is required. Whether risk can accumulate in the derivative position is dependent on the specialized terms of the contract and the triggers that are employed. In OTC derivatives, collateral triggers not only include changes in the value of the position but also changes in the credit quality of the counterparties. This approach differs from that of the exchanges, which do not generally adjust collateral based on creditworthiness, the portfolio risks of the counterparties, or their balance sheet. Pirrong (2009) finds these differences to be compelling and particularly relevant in the current market where accurate pricing of credit risk is critical. He posits that CCPs do not typically adjust collateral based on balance sheet risks due to the opacity of counterparty balance sheets and thus, "the prohibitively high cost to the CCP of evaluating, monitoring, and pricing these risks" (p. 39).

The process for exchange of collateral in the OTC markets is more decentralized than would be at the exchange and relies on the operational infrastructure of the OTC dealers or third parties they may employ (sometimes referred to as triparty collateral arrangements). Typically, collateral arrangements are agreed to and governed by the ISDA Credit Support Annex (CSA). The CSA specifies the terms for posting collateral and marking to market and also stipulates which counterparty would act as valuation agent and custodian or specifies a qualified institution to perform these functions. While subject to negotiation that would take into account the more customized features of the contract or the differences in credit quality of the counterparties, the mechanics of collateral management are similar to that of the exchanges.

The use of collateral as a means to ensure performance and provide a stock of cash or assets that could be drawn on in the event of default is common in the OTC markets, although less rigid than the margin practice used by exchanges. The ISDA Margin Survey (2008b) reports that dealers frequently require collateral at the inception of a transaction (akin to initial margin on a futures contract) and seek frequent revaluation of exposures in order to effect the transfer of gains and losses in the collateral accounts (in effect, transferring variation margin as a result of marking-to-market). As Exhibit 2.1 indicates, the growth of collateral agreements in OTC transactions has significantly increased, and the ISDA Margin Survey reports that 63 percent of OTC derivative transactions are secured by collateral. In addition, 65 percent of mark-to-market credit exposures are covered by collateral.

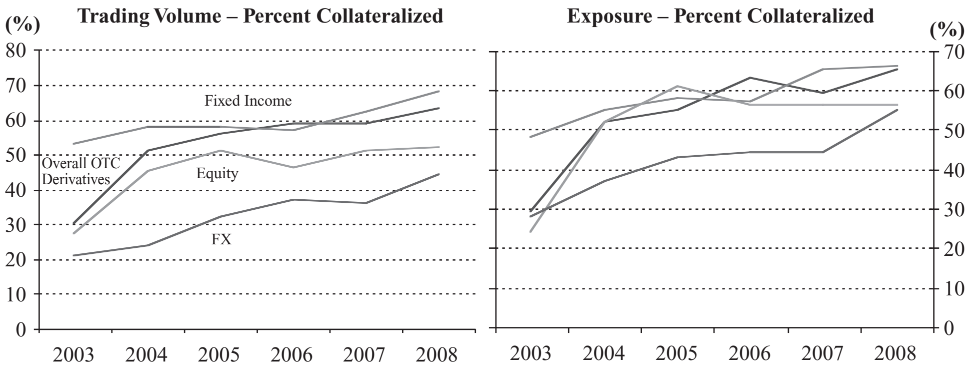


Exhibit 2.1 Use of Collateral in OTC Derivatives
Source: ISDA Margin Survey 2008.

The growing use of collateral and the use of triggers for further collateral calls have increased the linkages between firms but have also created an interdependence among large commercial counterparties. Just as with exchanges, the availability of collateral to cover losses in the event of default helps mitigate this risk. Pirrong (2008–9) notes that the linkages created by collateral arrangements in interdealer trade creates a sophisticated network for risk sharing that is less rigid than the formulaic approach used by exchanges. The practice of dealers to adjust their collateral requirements based on the position and balance sheet risks, and their tendency to negotiate the terms of collateral arrangements, should thus facilitate more efficient pricing of default risk than exists in the traditional exchange model.

In their CCP proposals for clearing CDS contracts, exchanges have recognized this drawback of the standardized model and have put forth various mechanisms for determining the collateral required in addition to expanded capital requirements. However, many have come to realize that in order to build a collateral pool sufficiently large enough to insure against default, such approaches may raise the requirements and inherent cost for market users of more customized and less liquid products. For these products, exchange officials have suggested margin levels for single-name CDS as high as 10 percent of notional value (*Risk*, 2009). Whether the costs exceed the marginal benefits that flow from the reduction in counterparty credit risk not already obtained in bilateral collateral arrangements remains to be explored, but the possibility exists that the increase in cost could curtail trading interest in CDS and make for a substantially less liquid market. In addition, as more standard products are forced onto a CCP, intermediaries will face reduced diversification of its collateral pool and potentially face increased risk and be discouraged from trading more customized products as well.

Netting and Novation in Exchange and OTC Markets

Derivatives users in the exchange and the OTC markets employ netting techniques across the pre- and posttrading functions. Prior to the trade, the technique of netting guides the assumption of risk in trading and intermediation in the front office of derivatives investors, brokers, and dealers. It is also frequently utilized in the middle office to inform the risk management activities of dealers, asset managers, and their principals, allowing the recognition of offsets and the monitoring of the risk of the traders, positions, particular strategies, and for the organization at large. But it is in the posttrade setting that netting delivers the greatest potential benefit by reducing actual obligations and payments resulting in effective reduction in settlement and counterparty default risk.

In posttrade processing, the practice of netting, or the reduction of payment and asset flows by the recognition of offsetting positions held among counterparties, is an important means by which exchanges and other financial networks have reduced risks associated with settlement (i.e., with respect to payments and deliveries associated with transactions). Until recently, OTC dealers primarily employed bilateral netting techniques in positions and payments. In the bilateral context, position netting is accomplished when two counterparties recognize and offset similar obligations they have to each other. For example, assume Party A has a contract obligating it to deliver 50 securities to Party B and Party B has a contract obligating it to deliver 100 of the same securities to Party A. If the counterparties agree to net their obligations, Party B would make a single transfer of 50 securities to Party A. As a result of netting, two transfers are reduced to one, lowering settlement risk. In addition, Party A is now exposed to nonperformance on only 50 securities, while Party B's exposure is reduced to zero, resulting in lower credit exposure for both counterparties (Culp 2004).

Position netting between counterparties in OTC derivatives is complicated by the fact that some privately negotiated transactions contain customized features that limit their fungibility. Contracts on more standard OTC derivatives are amenable for position netting, and more widespread application of position netting could lower delivery costs as well as reduce settlement risks. At the time of the publication of his book, Culp (2004) noted that of the four types of netting (position, payment, novation, and closeout) only closeout netting was commonly applied to OTC transactions. While netting has long been identified as a goal of dealers in the OTC markets, netting on a widespread basis has been hampered by the necessity of adopting and confirming bilateral netting agreements (e.g., through netting provisions in ISDA Master Agreements) and the prevalence of manual processes in the delivery and payment systems (ISDA, 2004).

Netting of payments for the same asset or in the same currency, referred to as payment netting, also reduces settlement risks. Daily bilateral netting of spot payments, in dollar and other currencies, including payments associated with the settlement of derivatives transactions, is processed through settlement banks such as the Clearing House Interbank Payments System and the SWIFT system. In addition, the CLS (Continuous Linked Settlement) Bank was formed by banks to facilitate bilateral and multilateral netting for foreign exchange transactions, and its potential for use for settlement of a variety of open derivatives transactions quickly became apparent as payments associated with settlement events increased (Bank for International Settlements 2007).

Multilateral netting reduces risk by aggregating the exposures of various bilateral counterparties arising from their derivatives obligations, calculating and recognizing offsets, and assigning the incremental exposure to each counterparty. Members agree to offset trades within the group, treating each other's contracts as perfect substitutes; original bilateral exposures are extinguished and multilateral net exposures reallocated among participants so that each has a calculated net position. In this manner, multilateral netting results in a decrease in the funds required to collateralize the position, a reduction in total payments and asset flows, and a reduction in credit exposure (Jackson and Manning 2006, p. 5).

The most comprehensive application of multilateral netting in the OTC markets came with the marriage of DTCC's Trade Information Warehouse for CDS transactions and the payment processing and settlement capabilities of the CLS Bank. The DTCC-CLS partnership has enabled the calculation, processing, and multilateral netting of the vast majority of CDS transactions in multiple currencies. In its first quarterly settlement of outstanding CDS transactions, the DTCC-CLS arrangement for netting CDS transactions resulted in 98 percent reduction of trading obligations requiring financial settlement, for a total aggregate reduction down from $14.3 billion to $288 million net in U.S. dollar terms.[7]

In the exchange model, the use of netting is facilitated by the use of standardized contracts and the insertion of the exchange clearinghouse as the CCP to all transactions; simultaneous with every transaction, the exchange CCP becomes the buyer to every seller and the seller to every buyer, such that a purchase of a position is seamlessly offset by a sale of the same position. This process, referred to as novation, enables counterparties to swiftly change or liquidate their position in the open market, without concern for whether the original counterparty would accommodate the transaction. This structure of organized derivatives exchanges "facilitates trade among strangers," encourages the collection of information, and increases liquidity (Telser and Higinbotham 1977).

In the CCP structure, the clearinghouse performs the netting calculation and dispatches the payments on a multilateral basis. The clearinghouse enters the transaction between the two counterparties, replacing them with two contracts between each counterparty and the CCP. Since multilateral netting as performed by the clearinghouse requires the posting of collateral and requires daily payments based on market values, CCP clearing not only results in measurable reductions in settlement and payment risks but can also reduce liquidity and credit exposure.

Exhibit 2.2 provides an illustration of the potential for decreased exposure as payment and asset flows are reduced through progressive netting arrangements. In this stylized example, counterparties A, B, and C assume obligations to trade three derivative contracts, X, Y, and Z. The positions held by counterparties A, B, and C are listed directly beneath each counterparty. In addition to transactions with each other, A, B, and C are assumed to hold contracts with other unspecified counterparties denoted by U, and the counterparties to each transaction are identified with superscripts A, B, C, and U. To capture the potential to liquidate positions entered today with offsetting positions entered in subsequent periods, a subscript of one or two denotes the time period the transaction is made. Each position in X, Y, and Z is preceded by a + if it is a long position and a – if a short position, and is listed in the boxes below each counterparty.

When the parties implement bilateral netting, obligations are reduced from an initial level of 52 outstanding positions to 16 open positions as participants are able to net positions with the same counterparty with whom they held the original position. The next level depicts multilateral netting in which payments associated with different assets can be netted against all the counterparties participating in

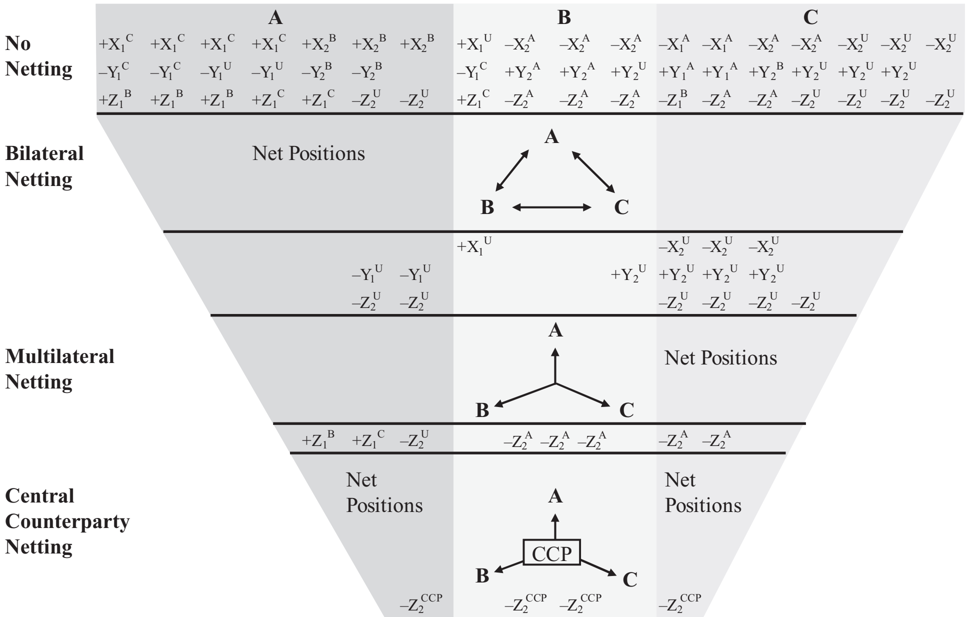


Exhibit 2.2 Decreasing Risk Exposure through Netting

the netting arrangement, decreasing the total exposure to four open positions. Finally, by establishing a CCP against which all open positions and payments can be netted regardless of the opposing counterparty or the time period in which they were entered, still more reductions are possible as counterparties are able to net across all open positions and payments created by the obligations, and total exposure is reduced to four positions. In the CCP case depicted at the bottom of the diagram, the only credit risk faced by each counterparty emanates from the possibility that the CCP itself could fail.

While this discussion is directed toward reduction in exposure that results from coordination of delivery and payment events (such as swap reset payments) and final settlements, netting techniques and CCP clearing are also viewed as mechanisms to mitigate risks associated with counterparty default. When contracts are cleared on an exchange, the outstanding contracts of the nondefaulting parties are not necessarily closed out since the clearinghouse is the counterparty on the other side of all transactions. While this lowers the immediate risk for the counterparties to the defaulting party, the risk does not disappear but rather is spread among the members of the clearinghouse. In the event of default, the exchange will utilize the traded collateral (margin) of the defaulting party to cover any losses. If these funds are not sufficient, the mutualized collateral is used to cover the loss and internalize the recovery cost.

OTC dealers have developed default provisions, referred to as closeout netting, in the ISDA Master Agreements and Credit Support Documents that govern the closeout of contracts in the event of counterparty default. In general, these provisions enable closeout of outstanding contracts, calling for a lump-sum payment based on market quotations or losses as specified in the contract. Contracts also specify how collateral held by the parties will be applied or distributed to cover any losses or gains.

Adoption of netting techniques that result in the novation of trades and effectively reduces outstanding obligations lowers credit risk as fewer positions need to be closed out in the event of default. Referred to as tear-ups, such novations result in compression, or a reduction in contracts outstanding. OTC market service providers, such as the TriOptima system for the affirmation and settlement of derivatives as well as the DTCC DerivServ application described earlier, have introduced applications to facilitate tear-ups on a bilateral basis. But as is evident from the analysis, the most significant risk reduction comes from novation on a multilateral basis. In their system for multilateral trade terminations, Creditex and Markit administer portfolio compression runs in CDS transactions that resulted in the tear-up of $1.036 trillion in single-name CDS trades in 2008.[8]

While these payment and settlement systems developed by OTC dealers have resulted in the netting of a significant proportion of CDS exposures, the calls for imposing an exchange solution requiring a CCP structure have dominated the discussion regarding the CDS market. The writing of default protection on collateralized debt obligations (CDOs), written on mortgage-backed and asset-backed securities, would prove to be perilous for some banks and insurers that suffered significant write-downs as default rates on those securities increased in 2007 and 2008. While a relatively small component of the CDS market, comprising less than 1 percent of all CDS registered with the Trade Information Warehouse at the DTCC, the losses on CDS written on mortgage-related securities have led many to conclude that CDS were a prominent source of losses that contributed to the credit crisis.[9]

As the credit crisis worsened, efforts to reduce sources of risk that were believed to have contributed to the crisis increased attention from financial regulators and banking authorities, who pressed for CCP clearing of all CDS transactions. OTC dealers formed a consortium to develop a clearing model that would plug into DTCC's Trade Information Warehouse, preserving the DTCC model for storing, transmitting, and confirming trade terms that had grown to include the vast majority of CDS transactions of the major dealers and institutions. Exchanges from the securities and futures business, eager to provide services to the vast OTC marketplace, introduced their own solutions even as dealers themselves sped up efforts to satisfy government prodding to introduce CCP clearing. Exchange offerings, such as the CME and the Citadel Investment Group's platform for clearing CDS transactions, CMDX, put forth an infrastructure adapted from the futures model, combining an execution system for effecting trades and confirming their terms with the full suite of their CCP clearing system. The ICE acquired The Clearing Corp. and Creditex, and formed a bank holding company, the ICE Trust, for the purpose of clearing CDS transactions. Working with CDS dealers, the ICE Trust has successfully enabled the clearing of the more standardized CDS, such as those on CDS indices.

Even as the efforts to clear CDS proceeded, the framework in place for the closeout or transfer of derivatives positions was thoroughly tested in the Lehman bankruptcy, a major broker and clearing member for all the major derivatives exchanges as well as a significant dealer in OTC derivatives. In its exchange-traded positions, Lehman was able to transfer more than 135,000 accounts, and customer positions were quickly transferred to solvent clearing members. Similarly, over $400 billion in notional exposure emanating from CDS contracts referencing Lehman was netted (down to eventually $5.2 billion actually changing hands) and closed out as a result of the final settlement price determined in the ISDA CDS auction protocol.[10] As noted by former regulator Andrea Corcoran (2008, p. 44), the framework in place for both exchange and OTC derivatives markets in a default event "continues to provide incentives to market participants to resolve market events and to give the market some comfort as to expected results."

As OTC markets have moved from bilateral credit risk management to multilateral mechanisms, in many respects adopting the most effective features of the exchange model that fit the OTC market, they still retain many of the informational characteristics of a private contracting market. As privately negotiated deals, for example, there is less publicly available information about the terms of the contracts, including their prices, the collateral, and therefore the aggregate risk exposure inherent in the outstanding contracts. The next section considers the different informational characteristics of exchange and OTC markets, and considers proposals to require additional regulatory supervision akin to that of the organized exchanges.

TRANSPARENCY AND INFORMATION IN THE EXCHANGE AND OTC MARKETPLACES

One of the primary functions of organized exchanges that became evident early in their history is that the prices that result from transactions in the market are regarded as benchmarks for transactions off the exchange. The assembly of trading interest and the transmission of order information regarding the prices at which customers are willing to buy and sell allow market participants to assess market conditions and others' expectations. As information arrives from various sources and is reflected in the bids, offers, and transaction prices, the exchange mechanism acts as an information aggregator, and this contributes to highly efficient price discovery. As the primary mechanism of price discovery, exchanges have attracted both the commercial interest and the ire of producers, millers, elevator operators, refiners, utilities, distributors, and consumers, all of whom look to the prices generated in the markets for guidance in pricing their transactions in the cash commodity.

Various OTC contracts utilize the prices generated in the exchange marketplace. This feature, sometimes referred to as price basing, is exemplary of the complementary nature of the exchange and OTC markets. However, different approaches used in pricing complex products, in addition to conditions of illiquidity that have resulted in a paucity of prices, have increased suspicion that OTC markets represent an obscure source of market, credit, and systemic risk. Even as effective mechanisms for risk management have been adopted and have become common to OTC derivatives and asset management operations, their implementation in the OTC market has been called into question due to a perceived lack of transparency and the different methods of pricing the assets available in the OTC setting.[11]

Because of their structure, exchanges are able to centralize information collection and dissemination, leading to a high degree of transparency about the volume of assets traded and their values. Customized contracts characteristic of OTC markets are less amenable to the exchange model, since individual terms of contracts can be quite diverse, making prices and information harder to aggregate. Prices may differ due to terms related to the contract, the quality of the asset and delivery characteristics, differences in the times entered, the prevailing liquidity considerations, and even differences in the credit quality of the counterparties. For the same reasons, efforts to clear OTC products also face the challenge of finding adequate benchmarks to set their value and accommodate the mark-to-market process.

OTC trading systems, the most successful of which is the ICE, are sometimes referred to by critics as "dark markets" due to the perception that they are less transparent than organized exchanges. But it is important to remember that these markets are primarily designed as wholesale markets, with users confined to large commercial entities and institutional investors. Since these markets are not open to retail investors and are limited to large and sophisticated entities, regulation adopted in 2000 provided them with an exemption from many of the disclosure and supervisory responsibilities applied to exchanges. While the tiered regulatory approach was initially heralded as a practical means to accommodate the different characteristics of a proprietary market and its users, critics of OTC markets have argued that the exemptions give them an unfair competitive advantage over exchanges, which are subject to direct regulation by the Commodity Futures Trading Commission (CFTC), and could attract manipulators and speculators bent on avoiding the watchful eye of the regulator.

Unlike the bilateral markets they supplant, proprietary OTC markets like the ICE are characterized by greater transparency and specialized provisions for regulatory intervention for fraud or manipulation. However, as energy and other commodity prices rose to unprecedented levels in 2008, legislators became concerned that these OTC markets were being used by large institutions to take speculative positions that were driving prices and increasing volatility. This elicited regulatory proposals and legislative actions to require these OTC markets to adopt exchange structures, including position limits and increased disclosure requirements common to organized exchange markets. In addition to seeking additional disclosure of OTC positions for its surveillance function, the CFTC proposed to develop and publish periodic information regarding OTC dealers' positions in the futures markets (CFTC Staff Report 2008, p. 6).

The CFTC already publishes on a weekly basis an aggregated version of trader activity known as the Commitments of Traders (COT) report that shows the amount of positions held by large traders in the markets as well as whether those positions are held by commercial or noncommercial participants. As the prices of energy and other commodities rose in 2007 and 2008, many began to call for the CFTC to provide a specific breakout with respect to the composition of participants in the COT reports, with particular focus on hedge funds, commodity index investors, and the OTC dealers who issue derivatives contracts based on commodity indexes and cover them in the futures markets. Investor interest in taking positions in these indexes, such as the Goldman Sachs Commodity Index Fund, usually through OTC commodity swaps, has increased significantly in recent years.

Given the increase in trading volume in commodities and the simultaneous increase in commodity prices, some have suggested that the increased participation of index investors and the swaps dealers that serve them has had a distorting effect on the markets and had increased volatility (Masters and White, 2008). However, far from confirming these assertions, empirical research conducted by the CFTC and other federal agencies, as well as rigorous studies by economists, has continually rejected the thesis that increased participation of swaps dealers, hedge funds, or other commodity index funds causes higher prices, finding instead that they actually have a stabilizing effect on the price process (Interagency Task Force on Commodity Markets 2008). Büyükşahin et al. (2008) find that in crude oil in particular, the increased participation of these traders has increased the integration of prices across the maturity term structure. Increased integration across contract market maturities makes the contracts more effective hedging vehicles and improves linkages across futures prices. Contrary to the conventional wisdom, the growth of the commodity index swaps market not only increased the volume of the futures transactions but also enhanced the quality of the markets overall.

Since commodity index traders take large long positions and roll them from one maturity cycle to the next in a very deliberate fashion, their net positions are generally and consistently long. The roll of index investors from the active contract month to the next maturity can, however, be more costly if their trades are anticipated. While breaking them out as a separate category in the COT report is advocated to increase the transparency of the market and the trading value of the COT data to others, it would also potentially expose OTC dealers to increased risk that they would be front-run by others in the market (in effect, speculators who would trade ahead of the roll of the contracts given the information they could extract from the COT position reports and other sources).

In sum, given the diversity of OTC markets and the customized nature of the contracts, efforts to require increased transparency of positions and to pose limits on OTC positions are fraught with three operational challenges and difficult public policy questions.

1. Given the diverse investing and hedging motives of OTC market participants, it is difficult to easily classify traders as commercials or noncommercials for the purposes of aggregating positions and enforcing position limits.
2. Even as the CFTC is in possession of the information on positions and motives of OTC dealers as a consequence of their surveillance function, it is not yet demonstrated that the publication of positions helps markets function better. On the contrary, such publication through the COT reports may harm OTC market participants by revealing their strategies and their risk.
3. While it is hard to deny the public policy appeal of lower energy prices, greater transparency, and reduced risk for financial institutions that pose systemic risks, policy makers must ensure that the evidence supporting the transformation of OTC derivatives markets exists and is robust enough to justify such an unprecedented level of intervention into private contract markets.

CONCLUSION

The importance of assessing risk in derivatives transactions, and in how markets for those instruments contribute to systemic risk, has attained much greater urgency in the current economic crisis. The possibility that a default of a large derivatives dealer could give rise to explicit losses for their counterparties and that this could create a kind of domino effect of losses and potential defaults leading to a systemic event has deeply concerned the markets, regulators, and policy makers. Yet there is much to be learned from structures of the exchange and OTC derivatives markets and from in the manner in which intermediaries, including clearing firms and OTC dealers, and market participants have responded.

It seems a perennial thesis put forth by observers that derivatives markets (or, as investor Warren Buffet refers to them, weapons of mass destruction) pose untold risk to users, are home to unbridled speculation, and are unchecked sources of systemic risk. Yet with respect to both exchange and OTC derivative markets, surprisingly little in the form of objective analysis or credible evidence has been put forth to support the claims. That derivatives are a means for risk transfer, and when used as hedging and risk management tools result in the *reduction* of risk, seems to have been lost in the effort to lay blame for the crisis. This is true for exchange and OTC derivatives contract markets and is demonstrated many times in the remaining chapters of this text.

Perhaps because OTC derivatives are commonly quantified by the notional amounts underlying the contracts, many assume that those amounts represent amounts at stake or the total risks posed by the OTC derivatives. In fact, a more accurate analysis would involve an assessment of the risks posed by the expected payment streams based on those amounts. So the $684 trillion notional figure for OTC contracts stated at the outset may give some cause for concern even with an appreciation for collateral, netting, and other mechanisms used by dealers to mitigate risk. Further work focused on the applicability and effect of the mechanisms used by OTC derivatives dealers needs to be undertaken to better inform market participants and regulators and to keep from throwing the proverbial baby out with the bathwater.

To give context to the notional numbers, CME Clearing reported that the trades it cleared accounted for $1.2 quadrillion in notional value in 2008 and that it is managing approximately $111 billion in collateral and approximately $39 trillion in open interest.[12] Given the concentration of shared default risk into a single entity, the exchange clearinghouse, the extent to which exchanges can assess credit risk and overcome information asymmetries regarding their counterparties and the contracts themselves is critically important.

Ultimately, the challenge of OTC derivatives markets is a consequence of information asymmetries common to private contracts. As Pirrong (2009) notes, such information asymmetries are present in part due to the complexity of the OTC products, the business of OTC dealers to design and market custom products, and the necessity that they develop proprietary and specialized skills in the valuation and trading of these products. But just as exchanges have attained economies of scale and scope in the execution and clearing of standardized contracts, Pirrong observes that OTC dealers have developed economies of scope in evaluating and pricing credit risk. Shifting OTC derivatives to exchanges or requiring them to be cleared via a CCP model will not necessarily reduce information asymmetries common to customized contracts or result in more optimal risk sharing, although adapting them to the exchange model may make them more costly, limiting the variety and utility of contracts offered in the OTC framework.

As noted, linkages between large financial institutions are created through exchange-traded and OTC contracts, branching out through the clearing members of the exchange and OTC dealers alike. Given a default by a significant firm such as Lehman Brothers, which in fact served in both capacities, both OTC markets and exchanges were forced to contend with the closeout and transfer of positions and, when necessary, share the risks according to their respective obligations. And both exchange and OTC derivatives markets, as discussed, appear to have performed well, though important lessons are revealed by their experience (Corcoran, 2008). When global policy makers consider regulatory reform and the role of derivatives markets throughout the crisis, it is hoped that they will do so mindful of the significant contribution of exchange and OTC derivatives markets to the quality of the markets and the economy as a whole.

ENDNOTES

1. *Liquidity risk* refers to potential losses from a lack of trading, which makes positions difficult to exit or offset and to value. Credit risk refers to losses from failures of counterparties to make required payments.
2. ICE, Form 10-K for Fiscal Year Ended December 31, 2008 (filed with the SEC February 10, 2009), pp. 6 and 57.
3. CRMPG I was constituted to study and recommend responses to problems that manifested in the Long Term Capital Management episode, while CRMPG III was resurrected to assess and make recommendations related to the financial market crisis of 2007 and 2008.
4. Estimate based on Globex volume (CME Group, “Growth of CME Globex Platform: A Retrospective,” August 10, 2008).
5. In the Commodity Exchange Act, Core Principle 15 provides that: “The board of trade shall establish and enforce rules to minimize conflicts of interest in the decision-making process of the contract market and establish a process for resolving such conflicts of interest” (CEA Sec. 5(d)(15), 7 U.S.C. Sec. 7(d)(15)).
6. “ELX Prepares to Enter U.S. Futures Market,” News Briefs, Futures Industry, January/February 2009, p. 18.
7. “DTCC and CLS Bank International Launch Central Settlement of OTC Credit Derivatives Trades,” DTCC Press Release, January 14, 2008.
8. “Markit and Creditex Tear Up $1.036 Trillion CDS Trades: Outstanding Single-Name CDS Trades Reduced Significantly Through Portfolio Compression,” Creditex and Markit News Release, November 24, 2008.
9. “DTCC Addresses Misconceptions About Credit Default Swaps,” DTCC Press Release, October 11, 2008.
10. “ISDA CEO Notes Success of Lehman Settlement, Address CDS Misperceptions,” ISDA Press Release, October 21, 2008. Substantial netting preceded the auction, and DTCC reported that of the $72 billion outstanding and registered in the Trade Information Warehouse on October 21, 2008, that was reduced to $5.2 billion.
11. For a discussion of efforts to increase transparency through disclosure in financial statements, and the controversy surrounding the valuation of derivatives and structured products, see Brown-Hruska and Satwah (2009).
12. Reported by CME Clearing at http://cmdx.com/clearing-overview.html, accessed February 26, 2008.

REFERENCES

Acworth, W. 2005. "The Tipping Point: OTC Energy Takes Off," *Futures Industry* (January/February): 18–24.

Bank for International Settlements. 2008. *BIS Quarterly Review* (December).

Bank for International Settlements Committee on Payment and Settlement Systems. 2007. *Progress in Reducing Foreign Exchange Settlement Risk* (July).

Brown-Hruska, S., and S. Satwah. 2009. "Financial Disclosure and SFAS 157: Seeking Transparency in a Perfect Storm," *Capital Markets Law Journal* 4, no. 2: 1–22 (April).

Büyükşahin, B., M. S. Haigh, J. H. Harris, J. A. Overdahl, and M. A. Robe. 2008. "Fundamentals, Trader Activity and Derivative Pricing," Working paper, December 4, 2008.

Carlton, D. W. 1984. "1984: Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures," *Journal of Futures Markets* 4, no. 3: 237–271.

"Clearing Single-Name CDS May Prove Uneconomical." 2009. *Risk*, January 23, 2009.

Commodity Futures Trading Commission. 2008. *Staff Report on Commodity Swap Dealers and Index Traders with Commission Recommendations* (September).

Corcoran, A., 2008. "The Importance of Certainty," *Futures Industry* (November/December): 42–45.

Counterparty Risk Management Policy Group II. 2005. *Toward Greater Financial Stability: A Private Sector Perspective*, July 25.

Culp, C. 2004. *Risk Transfer: Derivatives in Theory and Practice*. Hoboken, NJ: John Wiley & Sons, 2004.

Domowitz, I. 1995. "Electronic Derivatives Exchanges: Implicit Mergers, Network Externalities, and Standardization," *Quarterly Review of Economics and Finance* 35, no. 2 (Summer): 163–175.

Durkin, B., and E. Gogol. 2003. "The CBOT-CME Common Clearing Link," *Futures Industry* (May/June): 16–17.

Falvey, J. M., and A. N. Kleit. 2007. "Commodity Exchanges and Antitrust," *Berkeley Business Law Journal* 4, no. 1: 123–176.

Futures Industry Association. 2009 *Monthly Volume Report* (December).

Futures Industry Association. 2008. *Volume of Futures and Options Trading on US Futures Exchanges, 1968–2007*.

Grossman, S. J., and M. H. Miller. 1988. "Liquidity and Market Structure," *Journal of Finance* 43, no. 3 (July 1988): 617–633.

Harding, P. C. 2002. *Mastering the ISDA Master Agreement: A Practical Guide for Negotiation*. Harlow, U.K.: Pearson Education Limited.

Interagency Task Force of Commodity Markets. 2008. *Interim Report on Crude Oil* (July).

International Swaps and Derivatives Association. 2004. *Moving Forward: An Implementation Plan* (May). Available at http://www.sitelevel.com/quclk.cgi?rd=http://www.isda.org/press/pdf/Implementation-Plan.pdf&res=26482a362d0dfc0d&crid=26482a362d0dfc0d&pos=1&mr=50&qu=Moving Forward: An Implementation Plan.

International Swaps and Derivatives Association. 2008a. *ISDA Margin Survey 2008*.

International Swaps and Derivatives Association. 2008b. *ISDA 2008 Operations Benchmarking Survey*.

Jackson, J. P., and M. J. Manning. 2006. "Comparing the Pre-Settlement Risk Implications of Alternative Clearing Arrangements," Working paper (March).

Masters, M. W., and A. K. White. 2008. *The Accidental Hunt Brothers: How Institutional Investors Are Driving Up Food and Energy Prices*, July 31. Available at http://accidentalhuntbrothers.com/?cat=1 Pirrong, C. 2008. "Clearing Up Misconceptions on Clearing," *Regulation* (Fall): 22–28.

Pirrong, C. 2008–9. "The Clearinghouse Cure," *Regulation* (Winter): 44–51.

Pirrong, C. 2009. "The Economics of Clearing in Derivatives Markets: Netting, Asymmetric Information, and the Sharing of Default Risks through a Central Counterparty," Working paper, January 8.

Telser, L. G., and H. N. Higinbotham. 1977. "Organized Futures Markets: Costs and Benefits," *Journal of Political Economy* 85, no. 5: 969–1000.

United States Department of Justice. 2008. "Comments before the Department of the Treasury, in Response to Request for Comments on the Regulatory Structure Associated with Financial Institutions," January 31.

United States Department of Justice. 2007. "Statement on Its Decision to Close Its Investigation of Chicago Mercantile Exchange Holdings Inc.'s Acquisition of CBOT Holdings Inc., June 11.

ABOUT THE AUTHOR

Sharon Brown-Hruska is vice president in National Economic Research Associate's Securities and Finance Practice. Prior to joining NERA in July 2006, she was appointed commissioner (2002–2006) and acting chairman (2004–2005) of the U.S. Commodity Futures Trading Commission (CFTC). She served on the President's Working Group on Financial Markets and has testified before Congress on the function and regulation of the U.S. futures and options markets. Dr. Brown-Hruska has advised on ISDA documentation and protocols, and has evaluated risk management and clearing mechanisms for exchange and OTC derivatives markets. She was an assistant professor of finance at George Mason University (1998–2002) and at Tulane University (1995–1998). Her articles have appeared in *Capital Markets Law Journal*, *Barron's*, *Journal of Futures Markets*, *Regulation*, and *Review of Futures Markets*. She holds a PhD and MA in economics and a BA in economics and international studies from Virginia Tech in Blacksburg, Virginia.

CHAPTER 3

Speculation and Hedging

GREG KUSERK
Deputy Director for Market Surveillance, Commodity Futures Trading Commission

In a market-driven economy, it is a fact of life that participants face financial exposures and risks. The homeowner who heats his house with heating oil faces the risk that oil prices will rise in the winter; the farmer who plants corn in the spring faces the risk that prices will decline by the time he sells his harvest in the fall; and the bank that writes a fixed rate 30-year mortgage faces the risk that interest rates will rise, causing the value of the mortgage to fall. Markets, in a sense, turn everyone into speculators in that all people face market price risks on the economic exposures they carry as prices move up and down.

While agents in a market economy face financial risks, methods exist for them to shed or otherwise mitigate these risks. The homeowner can contract with a heating oil supplier to fix the future purchase price for oil. Farmers can, and often do, enter into forward contracts to deliver corn to local elevators at a predetermined price. Banks can lend at floating rates, which guarantee that they will always receive a current market rate on their loans. Although these methods exist for economic agents to bring certainty to their financial exposures, they can at times be a cumbersome or inconvenient way of achieving such certainty. For example, forward contracts with heating oil suppliers or grain elevators lock the homeowner or farmer into a contractual arrangement with a specific counterparty. Banks may find it difficult to sell floating rate mortgages to homeowners wary of taking on interest rate risks.

As an alternative to these methods of dealing with financial risks, derivatives contracts have been developed to allow economic agents to keep their underlying economic risk while entering into a second financial contract designed to offset that risk. This practice generally is referred to as hedging. In essence, a hedger enters into a derivative contract to shift the price risk he possesses to a counterparty who enters into the opposite side of the contract.

The Commodity Futures Trading Commission (CFTC, 2006) defines hedging as:

> *Taking a position in a futures market opposite to a position held in the cash market to minimize the risk of financial loss from an adverse price change; or a purchase or sale of futures as a temporary substitute for a cash transaction that will occur later. One can hedge either a long cash market position (e.g., one owns the cash commodity) or a short cash market positions (e.g., one plans on buying the cash commodity in the future).*

The opposite of hedging is referred to as speculation. The *CFTC Glossary* defines a speculator as "an individual who does not hedge, but who trades with the objective of achieving profits through the successful anticipation of price movements." For example, an investor having no physical market position or exposure to euros would be described as a speculator if she entered into a CME Euro FX futures contract. This is because the investor has taken on a financial risk that she previously did not possess.

While these general definitions are easy to comprehend and serve as a sound base for exploring the concepts of hedging and speculation, they tend to be transaction specific and ignore the so-called hedger's or speculator's motivations for entering into the transactions. As a result, they treat hedging and speculation as opposite sides of a risk management coin. In fact, hedging and speculation, in their purest sense, are more like endpoints on a continuum of risk management strategies.

As we will see in this chapter, some firms or individuals who enter into what can technically be described as hedges are really engaged in speculative behavior, while others who would be classified as speculators may be entering into transactions designed to reduce their overall financial risk. A true assessment of whether someone is speculating or hedging, therefore, requires an assessment of their motivations for entering into a derivatives position as well as the composition of the portfolio of assets and liabilities they concurrently hold.

HEDGING TRANSACTIONS

To understand whether someone is truly hedging, speculating, or engaging in some activity in between, it is useful to understand how one uses a derivatives contract to offset a cash market risk. As the CFTC definition implies, a hedge position is one that involves two positions having values that move in opposite directions. One of the positions is in the cash market and normally is held as a result of the ongoing operation of an enterprise. The other position is in a derivative, such as a futures or option contract, and is entered into for the purpose of hedging.

A classic example of a hedging transaction is the farmer trying to protect himself from a fall in the value of crops he is producing. Consider the risk a farmer faces when he plants corn in the spring. The farmer has committed a certain amount of resources in terms of seed, fertilizer, and time to producing the corn. If prices are low in the fall when the farmer goes to sell his corn, he faces the risk that he will not be able to cover the costs of his inputs. In the language of the CFTC definition, he faces financial losses from adverse price changes. This risk can be offset, however, through the use of a futures contract.

Suppose the farmer believes that he will be able to get \$2.90 per bushel at harvest for his corn. Referring to Exhibit 3.1, on May 1 the farmer consults the Chicago Board of Trade price for December delivery of corn and sees that the contract is currently selling for \$3.00 per bushel. He decides to sell 10 futures contracts to cover his expected production of 50,000 bushels. On December 1, the price of corn in his local area is \$2.65 per bushel, insufficient to cover his production costs. However, at the same time, the price of a futures contract has fallen to \$2.75 per bushel. Since the farmer initially sold the futures contracts at \$3.00 per bushel, he can now purchase them back at \$2.75 per bushel, thereby generating a profit of

Exhibit 3.1 Hedging Corn Production with CBOT Corn Futures: The Perfect Hedge

Date	Cash Position	Futures Position
May 1	Plant corn with an expected yield of 50,000 bushels and a price of $2.90/bu	Sell 10 December corn futures at $3.00/bu
December 1	Sell 50,000 bushels of corn at $2.65/bu	Buy 10 December corn futures at $2.75/bu
Result Cash sales price $2.65 Gain from futures $0.25 Net sale price $2.90		

$0.25 per bushel, offsetting the lower price on his physical corn. The supplemental value from the futures transactions ultimately allows the farmer to net a sale price of $2.90 per bushel, thereby covering his production costs.

Whether an individual or company is hedging agricultural commodities, energy commodities or financial assets, hedges using derivatives contracts, whether they are futures traded on an organized exchange or swaps traded off of an exchange, all function in generally the same fashion: They allow hedgers to hold a financial contract that fluctuates in value opposite that of the commodity or asset they are trying to hedge. In some cases, like the example of the farmer, the asset or commodity being hedged is already held by the hedger, who faces the financial risk that the commodity or asset being held will fall in value before it is sold. Under such circumstances, the hedger would sell a futures contract to hedge this exposure. Such hedges are referred to as short hedges because the hedger has sold, or "gone short," the futures contract.

In other situations, the hedger may have made a commitment to purchase a commodity or asset but not locked in a price. In those circumstances, the hedger faces the risk that prices will rise before the purchase is made. In this case, the hedger buys futures contracts to offset the exposure. If prices do rise, she can sell back her futures contracts at a higher price, thereby generating a profit to offset the higher purchase price of her physical commodities or assets. Hedgers using this strategy are said to have entered into a "long hedge."

Of course, hedges rarely if ever work out as neatly as in the example. In the example, the fall in the value of the futures price was sufficient to cover the difference between the farmer's cash price sale and his cost of production. But the local cash price at which the farmer could sell his corn may have been significantly lower than that realized in the example—for instance, $2.50 per bushel. In that case, the $0.25 earned on the futures contract when added to the cash sale would have netted the farmer only $2.75 per bushel, $0.15 below what he needed to cover his costs.

That hedges do not always work out perfectly, in the sense that a hedger cannot know with certainty that profits (or losses) on a futures position will exactly offset losses (gains) on the cash position, is due to a phenomenon called *basis risk*. *Basis* is defined as the difference between the cash price and the futures price. In the example, the basis on May 1 is –$0.10, as it is on December 1. If it is known that the basis will remain constant, then the hedger will have absolute confidence in his hedge since he knows with certainty that whatever changes occur in the value of what he is hedging will be perfectly offset by changes in the value of the futures positions. In this case, there is no basis risk. However, the value of a future basis is not known with certainty; therefore, hedgers face basis risk.

Basis risk can be thought of as the tendency for a hedge to perform better or worse than expected. If in the example a farmer's cash sales price drops to \$2.50 per bushel, the basis will have narrowed to –\$0.25 and the hedge performs worse than expected. Alternatively, if the cash sales price rises to \$2.80, the basis will have widened to \$0.05, and the hedge performs better than expected. In this case, the farmer's net sales price will be \$3.05.

The basis typically does not remain constant over time for a number of reasons, and hedgers therefore must face basis risk and deal with it. In the physical commodity markets, such as those for agricultural products, energy commodities, and metals, geographical and physical characteristics of the commodities themselves cause differences in prices and price movements to occur between that which is being hedged and that which can be delivered on the futures contract. For example, the NYMEX Light Sweet Crude Oil futures contract generally calls for the delivery of domestic U.S. crude oil with 0.42 percent sulfur by weight or less, not less than 37° API gravity or more than 42° API gravity. Delivery of the oil takes place in Cushing, Oklahoma. Given that most of the oil produced in the United States, or to an even wider extent globally, does not meet these specifications and delivery terms, the price of the oil being hedged will not be expected to move in exact correspondence with the price for the futures contract.

As with physical commodities, the prices of futures on financial assets, such as stocks and stock indexes, Treasury and other interest rate securities, and currencies, also do not move in exact correspondence with the assets being hedged. In the case of interest rate products, the assets being hedged, such as corporate bonds, commercial paper, or mortgages, may differ from the assets underlying the futures contracts due to differences in credit risk—for example, the credit risk of a corporate issuer versus the U.S. government—or in terms of differences in the interest rate duration of the debt. For stock portfolio managers, basis risk may exist due to differences in the composition of the stock portfolio being hedged versus the index underlying the futures contracts—that is, Standard & Poor's 500, Dow Jones, Nasdaq 100, and so on—versus a generic portfolio of stocks.

What is important to realize in pursuing a hedging program is that while a great deal of the price risk associated with holding a position in the physical market can be eliminated through hedging, basis risk will keep the hedger from eliminating all of it. To deal with basis risk, hedgers may adjust the size of the derivatives position held against a particular-size cash market position or make a decision to hedge or not depending on the current size of the basis compared to an expectation of where the basis may be when the hedge will be lifted. For example, if the corn farmer knows that the basis on December 1 is usually –\$0.10, but he sees that the basis on May 1 is +\$0.10, he may not place the hedge because he knows that if the basis narrows, as he expects, he will be worse off. This is shown in Exhibit 3.2.

In this case, the farmer realizes a net sale price of only \$2.70 for his corn rather than the \$2.90 he expected and needed to cover his cost of production. This shortfall can be traced to his loss on the futures position being larger than the gain on the

Exhibit 3.2 Hedging Corn Production with CBOT Corn Futures: Basis Risk

Date	Cash Position	Futures Position	Basis
May 1	Plant corn with an expected yield of 50,000 bushels and a price of \$2.90/bu	Sell 10 December corn futures at \$2.80/bu	+\$0.10
December 1	Sell 50,000 bushels of corn at \$3.00/bu	Buy 10 December corn futures at \$3.10/bu	−\$0.10
Result			
Cash sales price \$3.00			
Loss from futures \$0.30			
Net sale price \$2.70			

cash crop. Moreover, the fact that he fell 20 cents short of his expected sale price of \$2.90 matches his expectation that the basis would narrow by 20 cents.

Although the farmer in this case would have been better off not placing the hedge than placing it, he would still protect himself from even greater financial disaster by hedging than not. If, for example, the harvest time sale price is extremely low, the futures contract could still offer some protection even if the basis narrows. Exhibit 3.3 shows this situation. In this scenario, the basis still narrows by 20 cents, but the cash sale price at harvest is only \$2.60. In this case, the farmer is better off by having placed the hedge.

These examples illustrate that hedgers are never able to entirely rid themselves of the financial price risks they face. In the case where hedgers choose to place a hedge, they face basis risk. As a result, most hedgers are very aware of basis relationships and make their hedging decisions based on this knowledge. Working (1953) finds that most hedging is done based on expectations that the cash-futures price relationship—that is, the basis—changes over time in a predictable way. For this reason, it is often said that hedgers end up speculating on the basis. However, to the extent that the basis relationship between what is being hedged and the contracts being used to hedge is less volatile than the volatility of the commodity or asset being hedged, hedgers can reduce their overall financial risks by hedging.

Exhibit 3.3 Hedging Corn Production with CBOT Corn Futures: Basis Risk versus Price Prediction Error

Date	Cash Position	Futures Position	Basis
May 1	Plant corn with an expected yield of 50,000 bushels and a price of \$2.90/bu	Sell 10 December corn futures at \$2.80/bu	+\$0.10
December 1	Sell 50,000 bushels of corn at \$2.60/bu	Buy 10 December corn futures at \$2.70/bu	−\$0.10
Result			
Cash sales price \$2.60			
Gain from futures \$0.10			
Net sale price \$2.70			

SPECULATION

If the classic depiction of hedging is the farmer locking in a sale price for a crop, the classic speculator is the commodities trader who buys a pork bellies contract for no other reason than he believes prices will rise. Years ago, such speculators were limited to agricultural products or precious metals. Today, individuals wanting to chase profits have a panoply of choices available to them, ranging from physical commodities—agricultural products, metals and energy—to financial products—interest rate instruments, stocks, and foreign currencies. In today's markets, one can even take a view on temperature, rainfall, or even election results.

For the moment, as with the case of hedgers in the previous section, we will consider speculation in its purest form. That is, speculation is the taking on of a price risk for the simple purpose of trying to profit based on expectations of which way prices will move. Speculators who expect prices to rise will enter into long positions, while those believing they will fall short the market.

Speculators can be categorized in several different ways. One common way is to classify them by how they form their price expectations. Those relying on basic economic conditions to form expectations are referred to as fundamental traders. Traders in an alternate group, which form expectations based on analyses of price patterns and other market statistics, are called technical traders.

Fundamental traders operate on the premise that futures prices reflect the underlying conditions related to the supply and demand of commodities and the valuation of financial assets. The goal of the fundamental trader is first to identify the key economic conditions and variables that affect prices and second to observe changes in those conditions, it is hoped, before they are incorporated into the market price. For example, a fundamental trader interested in trading Eurodollar futures would be concerned with Fed policy, inflation rates, and other economic indicators that would signal upcoming changes to Eurodollar rates. A fundamental trader of physical commodities similarly would be concerned with factors that would increase or decrease the supply or demand for a commodity. If the trader can gain an edge in terms of gathering information on these factors and take a position in the market before the market has incorporated the information, she stands to gain from her efforts.

Technical traders operate on a premise virtually diametrically opposed to that of fundamental traders. These traders believe that knowledge of fundamental economic conditions offer traders little opportunity to profit in the market. While they may believe that markets reflect fundamental economic conditions, they also believe that markets, or at least the individuals participating in them, tend to repeat themselves or trade in ways that cause certain patterns to exist and that can serve as predictions of future price movements. The goal of technical traders, therefore, is to identify and correctly interpret the patterns they observe in the market and place trades accordingly.

The focus of technical traders is primarily on price patterns. Charting daily high, low, and closing prices, technical traders will seek out familiar patterns for signals that indicate that prices may be about to rise or fall. They may also examine moving averages to make their predictions. Finally, technical traders will also consider nonprice market statistics, such as volume and open interest or surveys of investor sentiment, to try to predict what actions participants may take in the market and their effect on future price moves. For example, a market in which the short open interest held by commercial participants is unusually high may be taken as a signal that prices will fall, on the assumption that few remaining commercials are left outside the market to take new short positions. Therefore, buying pressure would be expected to make itself felt in the market.

While fundamental and technical analysis has been in existence for years, the debate over which, if either, method offers greater chance for success rages on. Adherents to the efficient market hypothesis, such as Malkiel (2003a, 2003b) argue that there is little support for technical analysis (or for that matter fundamental analysis). They argue that markets are efficient and therefore that prices reflect all public information. Under this assumption, past prices, or patterns of prices, should not contain any information that would allow them to be used to predict future prices. Moreover, critics of technical analysis argue that to the extent useful price patterns did arise, market participants would quickly identify them and prevent them from reoccurring.

Tests of the efficacy of technical trading have produced mixed results. Articles by Lukac, Brorsen, and Irwin (1988), Brock, Lakonishok, and LeBaron (1992), and Osler (2000) suggest that technical analysis may produce useful information for predicting future price movements. However, subsequent research by Park and Irwin (2004) and Cooper and Gulen (2006) counter that these earlier studies suffer from various research problems, including data snooping,[1] *ex post* selection of trading rules, and difficulties in the estimation of risk and transaction costs—that is, whether a trader could generate excess profits for a given level of risk based on the trading signals produced by technical trading rules. Whether technical analysis works or not—and this will surely continue to be debated—what is clear is that it continues to be used by a significant portion of the speculative public. A simple Internet search of the term *technical analysis* will return literally thousands of sites offering advice or software or soliciting business to trade futures contracts using technical analysis.

Of course, just as there are those who raise issues with technical analysis, critics of fundamental analysis also exist. The main criticism of fundamental analysis is not whether markets react to fundamental market conditions but whether traders can use public information to predict future price movements. The efficient markets hypothesis maintains that market prices reflect all public information and that there is no predictive value left in the information that fundamental traders look to in forming their predictions. That is, markets react so quickly to new information that individuals assessing this information would have little opportunity to react. Shostak (1997) and Shiller (2000), however, argue that markets are much slower to react to new information than the efficient market hypothesis implies. They suggest that it takes time for individuals and the market to evaluate new information and incorporate it into prices and that there are at times psychological impacts that enter into the market to make prices at least somewhat predictable. As with technical analysis, the debate over the efficient market hypothesis and the value of fundamental analysis will likely continue for some time.

In addition to categorizing speculators by how they form price expectations, they can also be categorized by how they trade. Speculators who hold positions for very short time periods—typically less than a day—are referred to as day traders. Day traders seek to make profits from the many small price movements that occur during the day, using the great amount of leverage they can command on a position to amplify those movements. It is also a hallmark of day trading that traders begin and end the day with no position in the market. Because these traders are in and out of the market frequently, they tend to rely on technical analysis, which can generate trading signals more often than fundamental analysis, which is slower to generate a signal.

Another type of trader in the market is referred to as a position trader. The position trader is the opposite of the day trader, holding positions over longer periods of time, perhaps days or weeks. These traders are the market tortoises as compared to the day trading hares. They look for long-term trends in the markets and generally seek out higher profits per trade than day traders. Because they are interested in these longer-term trends, they are more likely to use fundamental analysis to try to identify basic changes in the underlying economic fundamentals that influence prices, but they may also rely on technical analysis.

Speculators also enter into positions based on the relative differences between prices for various contracts. These traders are referred to as spread traders or arbitragers. Futures markets offer a diverse assortment of contracts by underlying commodity and based on expiration dates. Spread traders may therefore enter into trades seeking to profit from the change in relative values between contracts with different expiration dates—that is, calendar spreads—or between contracts with different underlying commodities. Spread traders constantly monitor the prices of these various contracts looking for opportunities to take positions in contracts that they believe have gotten out of line with each other. For example, if a trader believes that the crack spread in the oil markets is too large and expects it to decline, she can enter into a short crack spread position, which is a purchase of 3 NYMEX crude oil contracts and sale of 2 gasoline contracts and 1 heating oil contract. Likewise if a trader believes that the nearby CME S&P 500 future is overvalued compared to a more distant expiration, he can sell the nearby contract and purchase the more distant.

FROM HEDGING TO SPECULATION

Recall that at the beginning of this chapter, hedging and speculation were depicted to be less like opposites and more like endpoints on a continuum of risk management strategies. At one end of the continuum, one can think of hedging as the situation where hedgers take a derivatives position with the expectation that they can perfectly offset any price exposure in a physical market position. At the other end are the pure speculators who enter into a derivatives position that exposes them to price risk that they otherwise do not hold.

In reality, neither of these two situations is likely to occur, for two reasons.

1. It would be nearly impossible to construct the perfect hedge that rids a hedger of all price risk due to the basis risk that affects virtually all hedges.
2. Hedgers tend to form expectations as to how prices in the physical and derivatives market will move, both independently and jointly.

Johnson (1960) observed that "Traders may well undertake hedging activities but these activities are not independent of expected price changes. A hedge may be lifted, a long position taken in the future inventories adjusted, all on the basis of price expectations." Blanco, Lehman and Shimoda (2005) also note that airlines often "hedge" fuel purchases based on expectations that prices will rise. As they point out, such hedges are really speculation as the airlines are trying to earn profits from predicting fuel prices. That is, they place hedges when they believe prices will rise and remain unhedged when they believe they will fall.

In addition to using derivatives to selectively place hedges, derivatives users employ a variety of other strategies that can be described as lying somewhere between pure hedging and pure speculation. Let us begin with the premise that hedging is the holding of a derivative position that offsets a cash market position in the same commodity. If a derivative contract does not exist for the particular commodity being hedged, the hedger must resort to something called cross-hedging. Such is the case with airlines that hedge jet fuel purchases. Because a jet fuel futures contract does not exists, airlines will take positions in heating oil or crude oil, which are closely though not perfectly related to the price of jet fuel. Such hedges will face greater basis risk than if a direct hedge using a jet fuel contract was available. Thus, to the extent that a hedger is a speculator in the basis, the cross-hedger is likely to be more of a speculator than someone who can directly hedge. Nonetheless, as long as the basis risk is less than the overall price risk, the hedge will still serve to lower the hedger's overall price risk.

As one shifts from hedging to cross-hedging, the issue of price correlation comes more into play as the hedger has to be concerned with whether the correlation between the hedged position and the derivative contract is sufficient to make the hedge perform well. The higher the correlation between the two, the better the hedge will perform since an opposite position in the derivatives position will be negatively correlated with the physical position. Of course, as the portfolio of risks a potential hedger holds grows—say from that of a farmer growing a single crop to a portfolio manager holding a diverse array of stocks and bonds—these correlations begin to be more important in a portfolio sense rather than on the one-to-one basis that a farmer or oil refiner might look at them. That is, it becomes more appropriate to evaluate the effect of the derivatives positions on the overall portfolio rather than attempting to tie each derivatives position to a particular transaction within the portfolio.

In the 1970s and 1980s, financial futures were introduced on interest rates and stock indexes. These instruments initially proved useful to portfolio managers wanting to hedge their portfolios in the traditional sense or to improve the liquidity of their portfolios—for example, when a manager wanted or needed to quickly change the exposure or investment mix of the portfolio. Today, however, many portfolio managers have taken a broader look at derivatives, including using derivatives on physical commodities as an asset class and means of improving the risk/return characteristics of their portfolios. What is interesting about this development is that while the inclusion of, say, agricultural futures contracts in a stock portfolio can hardly be described as hedging, their inclusion in the portfolio nonetheless may reduce the riskiness of the portfolio. So while including commodity futures in a portfolio may not be hedging, it certainly invokes the spirit of hedging.

Studies by Jensen, Johnson, and Mercer (2000), Erb and Harvey (2005), and Gorton and Rouwenhorst (2005) all show that the inclusion of commodity futures contracts in a portfolio can improve the performance of stock and bond portfolios. Gorton and Rouwenhorst attribute the improved performance to the effectiveness of futures in diversifying portfolios. They find that commodity futures tend to be negatively correlated with stocks and bonds in periods of unexpected inflation. They also find that futures diversify the cyclical variation found in stock and bond returns. Jensen, Johnson, and Mercer also find that return/risk optimization gives substantial weight to the inclusion of commodity futures in traditional portfolios. Moreover, they find that the inclusion of commodity futures during periods of expansive monetary policies plays an even greater role in improving the performance of these portfolios than in periods of restrictive monetary policy. Gorton and Rouwenhorst find that portfolios stand to benefit most when portfolio managers use tactical asset allocation strategies rather than long-only futures investment schemes.

In addition to futures being useful to portfolio managers in diversifying the assets they hold, futures have also been used to manage or enhance the liquidity of portfolios. Due to trading decisions or a need to increase or liquidate assets in a portfolio, portfolio managers may find themselves in a position of trying to either buy or sell large amounts of assets over short time periods. Under these circumstances, derivatives such as futures may offer a way to quickly add or subtract this exposure in a portfolio without having to transact in an illiquid market. Later as these portfolio managers purchase or sell the assets, they can reduce the futures position. While this practice may be viewed as hedging, given that the futures position is an offset to an underlying asset position or to a future purchase of an asset, it is also part of an overall speculative strategy to expose a portfolio to price risk. So as one moves from direct hedging, to cross-hedging, to portfolio management, at some point a line is crossed that distinguishes hedging from speculation. The exact point at which that occurs, however, may not be entirely clear, or perhaps even relevant.

INTERACTION BETWEEN HEDGERS AND SPECULATORS

While hedgers and speculators have different motivations for entering the derivatives markets, they are nonetheless very connected and indeed very dependent on each other in the marketplace. As discussed, hedgers come in two forms: those seeking to sell contracts to hedge future sales and those seeking to buy contracts to hedge future purchases. If markets relied solely on hedgers to trade in the market, two problems would inevitably arise: Hedgers with opposite hedging needs might not show up in the market at the same time, and there may be greater demand for hedging on one side of the market than the other.

Under either of the circumstances, hedgers are faced with the dilemma of either having to remain unhedged while waiting for a counterparty to arrive in the marketplace or making a large price concession in order to coax someone into entering a trade opposite of them. This situation is often referred to as a lack of immediacy or liquidity in terms of executing a contract. A special type of speculator, referred to a scalper, is the one who fills the void, providing immediacy in the market. Scalpers stand ready to purchase or sell contracts at any time. However, in order to take on the risk of holding a position for which they have no offsetting exposure, they will only be willing to sell at a price higher than they are willing to purchase at any point in time.

The spread between where a scalper is willing to buy and willing to sell is referred to as the bid-ask spread. Scalpers enter into these trades knowing that in a liquid market, they will quickly be able to sell everything they purchased, and vice versa. Scalpers generate their income by making small gains—perhaps a fraction of a cent—on a high number of trades. Many trades may even generate losses for the scalper, but the ultimate goal is to trade in high volume and have the gains outnumber the losses. Silber (1984) documents this behavior in scalpers, finding that scalpers provide an important source of liquidity to market orders entering the market. Thus, hedgers and other traders entering the market can quickly execute their trades due to the willingness of the scalpers to take the other side of contracts.

Although scalpers are an important source of liquidity for the markets, their impact on the market is short term. Ultimately it is the speculators who hold long-term positions that provide liquidity to hedgers in the market. Because scalpers liquidate positions as quickly as they enter into them, it is important that other speculators enter the market to take these positions. In addition, such speculators are at times necessary to handle any imbalances that exist between hedgers with opposite hedging needs. This role is filled by the day traders and, even more so, by position traders.

Invariably the mix between hedgers and speculators in a market will be driven by the characteristics of the industry underlying the futures contracts—that is, the demand for long and short hedging by participants in the industry—the perceived opportunity by speculators to profit in the market, the existence of alternative methods to hedge risks, the utility provided by a particular derivative contract to hedge or serve other risk management strategies, and so on. One source of information that is often used to gauge speculative and hedging use of the markets is the CFTC's Commitments of Traders (COT) reports. These reports are published weekly and show aggregate trader positions in actively traded futures and options markets. The reports contain information on large traders in the markets, indicating whether they are "commercial" or "noncommercial" traders.[2]

Generally academics and market analysts have considered the commercial open interest to be associated with hedging activity and noncommercial activity to be speculative. A detailed analysis of participation in the heating oil futures market by Ederington and Lee (2002) shows that while noncommercial open interest is almost certainly speculative in nature, a significant amount of activity by commercial traders may also be speculative. Nonetheless, they still found that 83 percent of average daily open interest and 75 percent of trading volume in their sample was accounted for by potential hedgers. Moreover, they found a rich diversity of trading styles among the 11 trader types they identified based on their line of business, with these styles being consistent with hedging and speculative strategies for these lines of business.

With this diversity in trading styles and hedging needs, the question often asked is: What is the impact on the markets by hedgers and speculators? A number of studies have attempted to answer this question. Ciner (2006) finds that in fact hedging dominates speculation in the NYMEX crude oil, heating oil, and unleaded gasoline futures contracts. These findings are consistent with those of Ederington and Lee (2000) showing that hedgers likely dominate the energy futures markets. Likewise, a study by Haigh, Hranaiova, and Overdahl (2007) also shows that speculators in the form of managed money traders tend to serve as liquidity providers in the markets and react to price changes as opposed to causing them. What these studies highlight is the interaction that takes place in the markets as speculators strive to profit from information and provide liquidity to the markets at the same time as hedgers seek to manage their price risk.

CONCLUSION

Hedging and speculation are often thought of as opposites. Hedging is the shedding of price risk, while speculation is the taking on of such risk. In practice, however, distinctions between the two activities are less clear. The continued development of portfolio and risk management theories and practices has blurred the distinction between pure hedging and pure speculation. Today we see hedgers that behave somewhat as speculators—as they choose to hedge or not hedge based on price expectations—and speculators behaving somewhat as hedgers—in the sense that they are using derivatives to reduce the overall risk of a portfolio. While early students of the futures markets, such as Working and Johnson, recognized that hedgers at times acted as speculators, it is only recently that, with the inclusion of commodity futures in investment portfolios, we have come to appreciate the subtle overlap between hedging and speculation. This appreciation suggests that the traditional definitions of speculation and hedging may have lost their relevance. In today's markets, practically all "hedgers" and "speculators" can be viewed as risk managers along the hedging-speculation continuum.

ENDNOTES

1. Data snooping can be described as the process of using the properties of a data set to influence the choice of a model or hypothesis. It often occurs when a researcher tests multiple hypotheses using the same data set until one is found to be statistically significant. Unfortunately, if enough hypotheses or models are tested, one will be found to be significant simply by chance. Thus, data snooping tends to overstate the statistical significance of models.
2. The CFTC requires futures commission merchants (FCMs) to report the end-of-the-day positions of traders holding positions of a certain specified size. The size differs by market, but generally is set so as to capture between 70 and 90 percent of the open interest in the market.

REFERENCES

Blanco, C., J. Lehman, and N. Shimoda. 2005. "Airlines Hedging Strategies: The Shareholder Value Perspective,"*Commodities Now* (June): 1–4.

Brock, W., J. Lakonishok, and B. LeBaron. 1992. "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," *Journal of Finance* 47, no. 5 (December): 1731–1764.

Ciner, C. 2006. "Hedging or Speculation in Derivative Markets: The Case of Energy Futures Contracts," *Applied Financial Economics Letters* 2, no. 3 (May): 189–192.

Cooper, M., and H. Gulen. 2006. "Is Time-Series-Based Predictability Evident in Real Time?" *Journal of Business* 79, no. 3 (May): 1263–1292.

Ederington, L., and J. H. Lee. 2002. "Who Trades Futures and How: Evidence from the Heating Oil Futures Markets," *Journal of Business* 75, no. 2 (April): 353–374.
Erb, C., and C. Harvey. 2005. "The Tactical and Strategic Value of Commodity Futures," NBER Working Paper No. 11222 (March).
Gorton, G., and G. Rouwenhorst. 2005. "Facts and Fantasies about Commodity Futures," Yale ICF Working Paper No. 04–20 (February).
Haigh, M. S., J. Hranaiova, and J. A. Overdahl. 2007. "Hedge Funds, Volatility, and Liquidity Provision in Energy Futures Markets," *Journal of Alternative Investments* 9, no. 4 (Spring) 10–38.
Jensen, G. R., R. R. Johnson, and J. M. Mercer. 2000. "Efficient Use of Commodity Futures in Diversified Portfolios," *Journal of Futures Markets* 20, no. 5 (May): 489–506.
Johnson, L. L. 1960. "The Theory of Hedging and Speculation in Commodity Futures," *Review of Economic Studies* 27, 3 (June): 139–151.
Lukac, L. P., B. W. Brorsen, and S. H. Irwin. 1988. "A Test of Futures Market Disequilibrium Using Twelve Different Technical Trading Systems," *Applied Economics* 20, no. 5 (May): 623–639.
Malkiel, B. G. 2003a. "The Efficient Market Hypothesis and Its Critics," *Journal of Economic Perspectives* 17, no. 1 (Winter): 59–82.
Malkiel, B. G. 2003b. *A Random Walk Down Wall Street: The Time Tested Strategy for Successful Investing*. New York: W. W. Norton.
Osler, K. 2000. "Support for Resistance: Technical Analysis and Intraday Exchange Rates," *Economic Policy Review* 6, no. 2 (July): 53–68.
Park, C. H., and S. H. Irwin. 2004. "The Profitability of Technical Analysis: A Review," AgMAS Project Research Report No. 2004–04 (October).
Shiller, R. J. 2000. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.
Shostak, F. 1997. "In Defense of Fundamental Analysis: A Critique of the Efficient Market Hypothesis," *Review of Austrian Economics* 10, no. 2 (September): 27–45.
Silber, W. L. 1984. "Marketmaker Behavior in an Auction Market: An Analysis of Scalpers in Futures Markets," *Journal of Finance* 29, no. 4 (September): 957–953.
U.S. Commodity Futures Trading Commission, 2006. *CFTC Glossary: A Guide to the Language of the Futures Industry* (Washington, D.C., July). Available online at www.cftc.gov/educationcenter/glossary/index.htm.
Working, H. 1953. "Futures Trading and Hedging," *American Economic Review* 43, no. 3 (June): 314–343.

ABOUT THE AUTHOR

Greg Kuserk is the deputy director for market surveillance at the Commodity Futures Trading Commission. He earned a BBA in business administration from the University of Notre Dame in 1980 and an MS in agricultural economics from the University of Delaware in 1983. Mr. Kuserk has been at the CFTC since 1987 and has served in several capacities as a research economist and policy advisor. He has provided technical assistance in the area of futures market regulation to exchanges and government regulators in Romania and Ethiopia and has authored a number of articles on the design and function of futures markets.

CHAPTER 4

The Social Functions of Financial Derivatives

CHRISTOPHER L. CULP
Senior Advisor, Compass Lexecon, and Adjunct Professor of Finance,
The University of Chicago Booth School of Business

Despite the long-standing concerns about their social costs, derivatives are among the oldest and most prevalent financial instruments in the global capital markets, having been used successfully by financial institutions, nonfinancial corporations, asset managers, and government-sponsored enterprises in their commercial, financial, and risk management activities. Many of the later chapters in this book review in more detail the various ways that firms can use derivatives constructively. Subsequent chapters also address some of the controversies surrounding derivatives, including the role they have played (if any) in recent financial crises and corporate scandals.

In this chapter, however, we explore a higher-level and more general topic than how derivatives have (or have not) impacted the fortunes of particular enterprises. Specifically, we consider here how the use of derivatives by some firms can benefit individuals and firms *apart from those directly participating in derivatives activity.*[1] These "social functions" of derivatives include:[2]

- Facilitating risk transfer.
- Serving as arenas for price discovery.
- Promoting the efficient allocation of resources to their most highly valued uses over time.
- Enhancing opportunities for investors to access alternative asset classes.
- Mitigating "underinvestment problems" by creating opportunities for asset-based financing.

This chapter discusses each of these social functions of derivatives in more detail, using historical and current examples where appropriate.[3]

The usual disclaimer applies: The opinions expressed herein do not necessarily reflect those of any organization with which the author is affiliated or their clients. Please address correspondence to the author to christopher.culp@chicagobooth.edu.

HEDGING AND RISK TRANSFER

Perhaps the archetypical social function provided by derivatives is risk transfer. Risk transfer is the process by which the adverse impacts of a risk are shifted from the shareholders of one firm to shareholders of one or more other firms (or individuals). The social benefits of risk transfer are thought to include reduced business failure rates, wider availability of consumer products in high-risk markets, increased opportunities for firms to invest in innovative but risky production technologies, and the redistribution of risks to those parties most willing and able to bear and manage them.

Firms routinely engage in risk transfer without using derivatives. A corporation that issues common stock, for example, is transferring the risks of its business to its stockholders. Similarly, vertical integration is a substitute for risk transfer—for example, a grain elevator can hedge against price increases and a farmer can hedge against price declines, or the grain elevator can acquire the farm (Carlton 1984). Buying and selling assets is also a form of risk transfer. A German chemical company concerned about the currency exposure on a factory it owns in Pakistan, for example, can sell the factory.

If a firm desires to transfer *specific* financial risks, however, the preceding methods of risk transfer may be overkill. The German chemical company that sells the Pakistani factory to eliminate currency risk also eliminates the revenues, costs, and strategic business role of the factory. Similarly, the grain elevator that buys the farm does solve its grain price risk management problem, but it also ends up having to own and operate a farm.

Derivatives can help firms selectively transfer risks, often at a relatively low cost and on flexible terms. Because they allow firms to select the particular risks being shifted, derivatives enable firms to manage risk surgically, freeing up managers to concentrate on those risks in which they have a perceived comparative informational advantage and to focus on running their businesses.

Consider, for example, lending and commercial banking. For many years, banks perceived themselves as having relatively better information about the credit risk of their borrowers than about the direction of future interest rates. Interest rate swaps and Eurodollar futures enabled banks to manage the interest rate risk of their banking books without changing their lending decisions. Credit derivatives and derivatives-based structured credit products, moreover, can also help banks fine-tune their risk management decisions.[4]

PRICE DISCOVERY

Price discovery describes the process by which trading in a market incorporates new information and market participants' expectations into asset prices. In a world without derivatives, the prices of traded securities and other assets presumably still would reflect market expectations. But thanks to the relatively low transaction costs and high liquidity of many derivatives markets, new information about assets is often reflected in derivatives prices first.

Reliable, public prices that reflect current information are essential in guiding the invisible hand for which the free price system is held in such high regard. Apart from promoting efficient resource allocation (as we discuss later), price discovery also assists firms in other ways. The term structure of futures prices, for example, is regarded for some assets as a good estimate of expected future spot prices. By looking at the term structure, firms can forecast their revenues and costs using market-wide information. Indeed, Roll (1984) observed that orange juice futures prices actually provide better information about weather forecasts than weather forecasts themselves.

Numerous commercial contracts, moreover, are negotiated by reference to corresponding derivatives markets. This is sometimes implicit—for example, a grain elevator that consults current futures prices before entering into a physical delivery contract. But often the relation between cash market prices and derivatives prices is explicit. An interesting example of this interdependence was documented by Kuserk and Locke (1994). In 1991, a tunnel system underneath the Chicago Loop flooded, leading to the temporary closure of the Chicago Board of Trade. During this time, grain elevators actually pulled down price quotes to farmers and did not repost those quotes until the futures markets reopened.

Options markets are also aggregators and providers of information. Market expectations of future price movements, for example, can be extracted from observed traded option prices and the volatility surface.[5] Banz and Miller (1978) show, for example, how this information can be used to guide investment and capital budgeting decisions by nonfinancial firms.

Price Discovery, Commoditization, and Market Structure

Commoditization is the process by which bilaterally negotiated, customized contracts evolve toward organized financial markets. Historically, price discovery has been associated mainly with commoditized derivatives such as futures. Indeed, futures exchanges generate significant revenues from sales of their price feeds to data vendors (Mulherin, Netter, and Overdahl 1991).

Not all customized contracts, however, evolve into standardized financial instruments traded in transparent markets. And those that do often spawn further evolutionary changes in custom, off-exchange contracts. Innovation that begins with customization thus evolves into standardization, which in turn begets further off-exchange innovation (Merton 1992).

Although organized futures exchanges remain arenas for price discovery in many markets, the concept and function of exchanges have changed significantly in recent years. Distinctions between exchange-traded and off-exchange derivatives have blurred, and the convergence of these markets has made it harder to draw clean lines among price discovery, transparency, and market structure.

Consider, for example, the market for Eurodollar derivatives (based on the 90-day London Interbank Offered Rate, or LIBOR). In the late 1980s and early 1990s, Eurodollar futures provided price discovery for short-term interbank funding markets. Fixed-for-floating interest rate swaps—also based on LIBOR—were growing rapidly at that time but were still mainly customized, opaque transactions. Over the next decade, however, interest rate swaps commoditized (while still remaining off-exchange). Bid-ask spreads on plain vanilla swaps tightened, and plain vanilla swap rates became more readily available from data vendors. Today, whether Eurodollar futures or swaps provide price discovery for interbank markets is no longer clear.

INTERTEMPORAL RESOURCE ALLOCATION

A significant social benefit of derivatives—forward and futures contracts, in particular[6]—is the role they play in rationing scarcity over time in the underlying assets on which they are based.

Forward Contracts as Synthetic Storage

Entering into a forward purchase agreement is known as synthetic storage because it is economically equivalent to the purchase and storage of the underlying asset. The forward purchase price is determined in equilibrium to ensure this is so. To see how this works, consider a firm that wants to own one unit of an asset (e.g., a bushel of wheat, a share of stock, a bond, a gold bullion bar) in three months. The firm can, of course, just wait and buy the asset in three months at its then-current spot price. Or the firm could buy the asset now and hold it for three months. Specifically, the firm borrows enough cash to buy the asset now and then holds it for three months. Over the intervening three months, the firm receives any cash distributions paid to asset owners (as well as any intangible benefits of having the asset on hand) but also bears any storage costs. And, of course, the firm must repay principal and interest on the cash loan at the end of the three months.

Alternatively, the firm might have entered into a forward purchase agreement to buy the asset three months later at a fixed price negotiated today. In the absence of arbitrage, the forward price of the asset to be delivered in three months should be equal to the current spot price plus the net cost of carrying the asset over three months (i.e., interest plus physical storage costs less the benefits of holding the asset).

If this "cost of carry" relation between forward and spot prices does not hold, at least some firms may be able to earn riskless profits by exploiting the deviation. Suppose the current price of an asset is \$100 and the three-month cost of carry is \$2 but the actual quoted three-month forward price is \$105. A firm could buy the asset and store it for three months for a total cost of \$102 and simultaneously sell the asset for delivery in three months for \$105, thus generating a riskless profit of \$3. This puts downward pressure on the forward price and upward pressure on the spot price. The process continues until the difference between the forward price, and its fair value is no greater than the transaction costs of the arbitrage.[7]

The cost of carry relation is an equilibrium condition that presumes a perfect capital market. Institutional frictions (e.g., transaction costs, liquidity constraints, restrictions on short sales, etc.) can drive a wedge between true prices and their fair values to the extent they inhibit arbitrage. During the stock market crash of October 1987, for example, operational problems (e.g., slow printers and systems) at the New York Stock Exchange made stock index arbitrage impossible for a time, causing the cash and futures markets to disconnect.[8] But setting aside such exceptions, the cost of carry relation between forward and spot prices tends to be reliable.[9]

Commodity Interest Rates

The relation between the price of an asset for immediate delivery and the price of that asset for future delivery characterizes an implicit commodity interest rate that guides the allocation of resources to their most highly valued uses over time (Keynes 1930; Sraffa 1932). On any date t, we can express the commodity interest rate prevailing through date T in this way:

$$\text{Commodity Interest Rate} = \frac{S(t) - F(t,\ T)}{S(t)}$$

where $F(t,T) =$ time t forward price of the asset for delivery on date T
$S(t) =$ current spot price

This commodity interest rate is equal to the marginal benefit less the marginal cost (both interest and storage) of holding the asset from t to T.[10]

The benefit of owning financial assets over time includes cash distributions such as stock dividends or bond coupons. But for commodities, the benefit of holding the asset is a "convenience yield" that reflects the implicit benefit of physical storage for firms that need access to the actual commodity in order to avoid costly production disruptions.[11] When inventories are high, plenty of the asset is available to assure producers and intermediaries that a stock-out will not occur. The convenience yield is small, and the price of the asset for future delivery is above the spot price by just enough to compensate asset holders for interest and storage costs. When the term structure of futures prices is positively sloped—called a carry or contango market—commodity interest rates are negative; a borrower of the commodity would pay more in storage and financing costs than it would receive in benefits from actual asset ownership. Firms thus have no particular economic incentive to take assets out of storage and sell them or loan them out today.

As inventories shrink, however, the marginal benefit of owning a unit of the underlying asset rises. The spot price thus goes up relative to prices for future delivery, resulting in a negatively sloped term structure of forward prices known as an inverted market or a market in backwardation. In an inverted market, current supply is low relative to demand, and firms will be willing to pay a positive commodity interest rate to get their hands on the physical asset today. Positive commodity interest rates in effect penalize firms that leave physical assets in storage for future delivery instead of bringing them out of inventory into the current, tight market.

Exhibit 4.1 illustrates this intertemporal supply rationing feature of derivatives using crude oil data from 1985 to 2007. The graph shows year-over-year percentage changes in total U.S. crude inventories (excluding the Special Petroleum Reserve) versus the three-month commodity interest rate based on Nymex futures prices. For most of the period, when commodity interest rates rise, oil inventories are drawn down.[12]

ASSET FINANCE

Firms that face rising deadweight external borrowing costs as a result of information asymmetries, credit constraints, and other market frictions sometimes must forgo positive net present value investments because of their limited access to affordable external financing (Froot, Scharfstein, and Stein 1993). Such so-called underinvestment problems can reduce capital formation and artificially depress

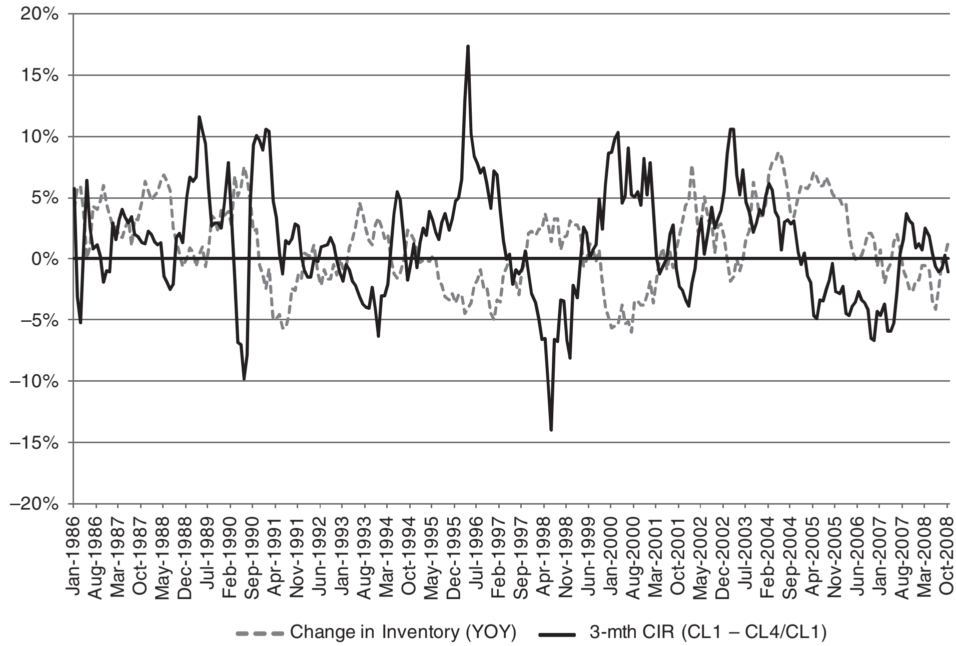


Exhibit 4.1 West Texas Intermediate Crude Inventories versus Commodity Interest Rates, 1986–2008

real investment activity. Derivatives can help mitigate underinvestment problems by enabling at least some firms to engage in asset-based financings.

Commodities Lending

As we saw in the previous section, forwards and futures facilitate intertemporal supply rationing by defining implicit commodity interest rates. In some markets, however, the borrowing and lending of assets is explicit, and the commodity interest rate is an observable market price.

Gold mines, for example, can borrow physical bullion from central banks and sell the gold spot to obtain immediate financing, repaying these gold loans later with gold from their own mines. The "gold lease rate" is the interest rate paid on a gold loan—for example, an annual gold lease rate of 5 percent on a 100-ounce gold loan means that 105 ounces of gold must be repaid to the gold lender in a year. The gold forward offered rate (i.e., GOFO) is generally above the spot price, which means LIBOR is at a premium to gold lease rates most of the time.[13] Exhibit 4.2 illustrates this. So, for mines facing high external financing costs, gold loans can provide a relatively cheap source of funds to finance capital investments.[14]

Commodity loans such as those we see in the gold market are hardly a recent innovation. On the contrary, they date back at least to Babylonia in 1900 B.C. to 1600 B.C. Like many other activities of the Mesopotamian era, banking in ancient Babylon was predominantly a religious practice centered around sanctuaries and temples (Jastrow 1911). The Temple of Šamaš (Ebabbara or Bit-Uri) at Sippar—a

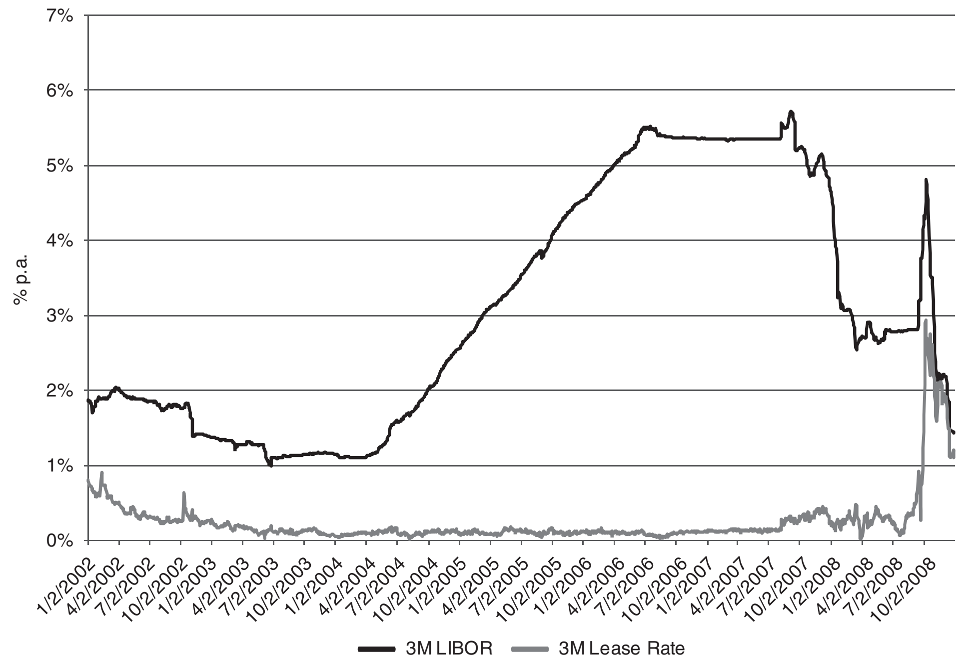


Exhibit 4.2 Three-Month Gold Lease Rates versus LIBOR, 2002–2008

shrine to Šamaš, the sun god and lord of justice and righteousness—was the dominant banking institution for much of the period (Bromberg 1942).

When the grain and livestock received by the Temple of Šamaš from tithes and offerings began to exceed storage capacity, the temple began lending grain to farmers who needed it to feed their families and workers and to sell for cash on the spot market to finance their operating cash flows. For example, the next quotation is translated from a loan tablet documenting a grain loan from the Temple of Šamaš to Minutum around the time of Sin-muballit, the father of Hammurabi:

> *10* gur *of grain—at the rate of 1/5th* gur *per* gur*—Minutum has borrowed from Šamaš. At harvest time he shall return it. (Bromberg 1942, p. 80)*

In other words, Minutum was obliged to return whatever grain he borrowed plus 20 percent to the Temple of Šamaš. Notice that this grain lending rate was quoted in quantity terms just as gold lease rates are today.[15]

Project Finance

A prepaid forward contract is a forward contract in which the seller agrees to deliver an asset in the future in return for receiving a payment from the buyer up front. Especially for firms with hard assets but limited access to unsecured credit, prepaid forwards can be an efficient method of monetizing those assets for financing purposes.[16]

Consider, for example, a developing country with below-investment-grade credit but a large supply of proven oil reserves. The country can sell its oil for future delivery using prepaid forwards, thereby generating cash up front that it may need to bring its oil fields into operation. Examples of such structures abound in modern structured project finance and are especially popular with credit-constrained borrowers.[17]

The use of prepaid forwards in commodity finance was prevalent as early as the sixth century in Islamic countries. Two principles of *shariah* are its prohibitions against usury and entering into transactions involving *gharar*, or risk. An important exception is called *salam*, defined as a sale in which specific goods are promised to a buyer for future delivery in exchange for a price paid in full up front. If a farmer needed cash (e.g., to pay his workers) before his crop came in, *salam* permitted the farmer to sell some or all of the crop before the harvest in return for a commitment to deliver grain in the future.[18]

Trade Finance

Derivatives also play an important role of trade finance. A Brazilian farmer selling coffee to Nestlé in Switzerland, for example, might go to a bank and obtain trade credit to finance his coffee exports. The bank likely will require the coffee to be pledged as collateral for the loan, along with a requirement that the farmer use derivatives or other means to lock in the value of the coffee collateral. Alternatively, the farmer might obtain pre-export financing and lock in his coffee sale price using a prepaid forward sale agreement with Nestlé.

One of the earliest examples of the use of derivatives in trade finance comes from medieval Europe, where the Medici Bank was operating very much like swap dealers today.[19] International trade in Europe up to the mid-1300s was conducted mainly at periodic regional fairs where merchants from various places came together in one location and exchanged their wares. This was problematic for merchants in two respects.

1. They had to finance their initial purchase of goods in their home currency.
2. Because correspondent and foreign bank branches did not maintain funds outside their home locations until the 15th century, merchants had little choice but to convert their foreign currency revenues back into their home currency. A Florentine merchant wishing to sell goods in Champagne, for example, thus needed to raise florins to finance his initial purchase of goods and would then later receive Provins money from the sale of those goods at the Fair of Champagne. (Face 1958)

In the fourteenth century, however, the church defined usurious lending as charging interest on a money loan in which the only risk to the lender was credit risk. Trade credit facilities like the bank loan to the Brazilian farmer in our earlier example thus were prohibited. But if the transaction was also subject to exchange rate risk, church usury prohibitions did not apply.

So the Medici Bank began offering a contract called a cambium that combined foreign exchange and commercial lending in a single transaction. An ordinary cambium involved a loan to the merchant in his local currency in exchange for the subsequent repayment of funds in a different currency and location at a prespecified price—that is, a currency forward. Because both the merchant and bank were exposed to exchange rate risk, the church did not view cambiums as usurious and thus permitted them.

In the late fourteenth century, the Fair of Champagne declined in popularity and trade migrated to fairs in larger cities, such as London and Bruges, where countinghouses arose at which merchants maintained foreign currency balances. In response to competition from those countinghouses, the banking community replaced ordinary cambiums (which were custom, bilateral transactions) with standardized bearer certificates called *lettera di pagamento* or *lettera di cambio* (bills of exchange), which were prepaid currency forwards by another name. Organized markets for trading bills of exchange eventually emerged, with the Medici Bank as the dominant market maker. As the number of correspondents, branches, and agents of the Medici Bank grew in the fourteenth and fifteenth centuries, published quotes by the Medici Bank became *the* central source of information about foreign exchange rates (De Roover 1963, p. 122).

Financial Asset Inventory Management

Derivatives also can be used for asset finance by financial intermediaries. Although perhaps not an example of how credit-constrained firms can use derivatives and asset finance to mitigate underinvestment problems, derivatives do enable securities dealers to liquefy their securities portfolios and finance their securities inventories at a lower cost than if they had to borrow unsecured. This can reduce their costs of liquidity provision and market making, thereby promoting more efficient and liquid underlying securities markets.

A government bond dealer, for example, can finance a bond position by selling the bond and agreeing to repurchase it later at a higher price. The difference between the two prices—the repo rate—represents the interest rate the dealer pays to its counterparty for a cash loan collateralized with the underlying bond. This interest rate is generally below commercial borrowing rates such as LIBOR and is the closest a private corporation can get to the Treasury borrowing rate. A repo is, of course, just a forward contract combined with a sale of the security in the spot market—that is, a synthetic loan of the security.

In a stock loan, the stock borrower (e.g., a short seller who needs to borrow the stock to honor a security sale commitment) enters into an agreement with a stock lender to acquire shares in return for an obligation to return those shares whenever the stock lender wants. The borrower posts collateral at the time the shares are borrowed. When the shares are returned, the borrower receives back the collateral plus a rebate of some or all of the interest. Shares in higher demand command lower rebates, and hot issues may have negative rebates, in which case the stock borrower must make an additional payment to the lender above and beyond the forgone interest on the collateral.[20] Stock loans thus are also types of repos (albeit undated) in which the stock lender sells shares for cash and later repurchases them for a total net cost based on the rebate rate.

SYNTHETIC ASSET ALLOCATION

Investors often find derivatives appealing because they provide low-cost alternatives for investing in asset classes that, absent derivatives, are prohibitively costly or operationally difficult to hold otherwise. If these new asset classes have low or negative correlations with existing major asset classes, adding them to the investment opportunity set enables investors to achieve their target expected returns with less risk. Investors can access these opportunities by using derivatives directly (possibly on a fully collateralized basis) or by investing in funds that use derivatives, such as managed futures funds, certain hedge funds and structured investment vehicles, and collateralized commodity obligations.

Gorton and Rouwenhorst (2006) show, for example, that commodities have similar Sharpe ratios (i.e., average return per unit of risk) to equities but returns that are negatively correlated with stocks and bonds. Not surprisingly, the growth in managed commodities funds seeking to engage in synthetic asset allocation into commodities has been significant over the past decade.

Some have also characterized volatility as an asset class. In equities, for example, volatility is negatively correlated with stock index returns.[21] Option spread trades, such as straddles and strangles, allow traders to take directionless positions on volatility. More recently, volatility derivatives, such as variance swaps, have become a popular tool for investing in volatility as an asset class.

Apart from these examples of synthetic asset allocation, numerous hedge funds and structured product vehicles make use of a wide range of active strategies involving derivatives for the purpose of enhancing returns. Whether such strategies actually expand the efficient portfolio opportunity set (as opposed to simply repackaging existing opportunities or adding leverage) is an empirical question.

Derivatives and Public Policy

Empirical financial economics and econometrics cannot tell us definitively if derivatives have been a net benefit to society in the nearly 4,000 years they have been around. Nevertheless, the reasons to believe that derivatives have done more good for society than harm are compelling.

As financial entrepreneurs continue to develop new derivatives, some surely will come under scrutiny, especially if such products are associated with headline-making losses—as some inevitably will be. The temptation of derivatives critics in such situations will be to respond with political proposals to restrict or deter some of these ostensibly dangerous new derivatives products. In the wake of the credit crisis, for example, criticisms of derivatives (especially credit derivatives) and demands for new regulations have been as numerous as they have been vigorous. But that does not necessarily mean those criticisms are well founded or that the proposed regulations are the appropriate response.[22]

Public policy decisions concerning the risks of innovative financial structures should be made with prudent deliberation and based on the empirical evidence. The risks to society of restricting unproven but controversial financial innovations, after all, may well exceed any risks of the innovations themselves. Smith (2003) makes this point compellingly:

> *Civilization can be seen as the gradual evolution of ever more creative risk management—from the family and private property to derivatives and structured financing arrangements. The goal is to permit an ever greater scope for the prudent assumption of risk. Because knowledge is dispersed, only that expanded scope offers any hope of fully using the varied skills of all the peoples of this planet. Civilization is the story of the advances and retreats of such prudent risk management expansions.*

> *Civilization makes it possible to better manage risks in the financial, technological, and social fields. Indeed, a reasonable metric for assessing the level of civilization is mankind's success in evolving institutions that permit an ever-larger scope of prudent risk taking. Prudence is best defined as a careful calculation of the risks of change versus the risks of stagnation—and the development of institutions that encourage that careful balancing. (p. 266–267)*

An appropriate metric of civilization is our ability to manage the risk of innovation—business, financial, technological, and social innovation alike. Indeed, a reasonable metric for assessing the level of civilization *is* mankind's success in evolving institutions that permit an ever larger scope of prudent risk taking. History is the story of mankind's slow stride from tribal collectivism to modern individualism, from poverty to affluence. Prudence is best defined as a careful calculation of the risks of change versus the risks of stagnation—and the development of institutions that encourage that careful balancing.

ENDNOTES

1. Despite the theoretical and empirical problems with estimating "social welfare" (see, e.g., Demsetz 1969), a lot of empirical work has been done on derivatives, much of which aims to address more specific and tractable questions than whether derivatives are "socially good" or "socially bad." Nevertheless, this chapter makes no claim to be a literature survey. Certain references to the literature are provided when appropriate, but these references are not intended to be and are not exhaustive.
2. Not all derivatives perform all of these functions all of the time.
3. For a comprehensive review of the historical evolution of derivatives, see Swan (2000).
4. Some contend that by facilitating risk transfer, credit derivatives were in part to blame for the subprime crisis. An article in *Fortune* magazine, for example, claimed that "by ostensibly providing 'insurance' on risky mortgage bonds, [credit derivatives] encouraged and enabled reckless behavior during the housing bubble" (Varchaver and Benner 2008). It remains to be seen whether the empirical evidence supports these (and numerous other recent) criticisms of credit derivatives. Addressing these criticisms, moreover, is beyond the scope of this chapter.
5. See, e.g., Jackwerth and Rubinstein (1996).
6. Despite some important institutional differences, we treat futures and forwards as economically equivalent in this discussion for expositional simplicity.
7. Not all firms face the same interest and storage costs or derive the same benefit from asset ownership. The cost of carry reflected in forward prices in equilibrium is the marginal cost of carry of the marginal market participant. In other words, not all firms will be able to exploit this arbitrage opportunity. But some will, and that is enough for the adjustment process to work as described.
8. See, e.g., Furbush (1989), Gammill and Marsh (1989), Harris (1989), Kleidon (1992), and Kleidon and Whaley (1992).
9. See, e.g., Fama and French (1987, 1988), Ng and Pirrong (1992) Ng, V. K., and C. Pirrong. "Fundamentals and Volatility: Storage, Spreads, and the Dynamics of Metals Prices." Journal of Business 67(2) (April 1994), 203–230. Stoll and Whaley (1990), Telser (1958), and Working (1948, 1949).
10. In this formulation, the benefits of asset ownership and the costs of physical storage would be expressed as a percentage of the spot price.
11. See, e.g., Williams (1986) and Working (1948, 1949).

12. Although positive commodity interest rates do penalize firms for storage, the converse is not usually true; negative commodity interest rates do not necessarily induce firms to store. In a "full carry" market, prices for future delivery exceed current prices by just enough to compensate for the interest and physical costs of storage, thus making firms indifferent between actual and synthetic storage. In recent times, however, the oil market has been in what some call "super-contango" in which prices of crude for future delivery exceed current spot prices by more than interest and storage costs, indicating a negative convenience yield and a positive return to storage. Not surprisingly, the tank farms in Cushing, Oklahoma—the delivery point for many crude derivatives—have been virtually at capacity during much of this period. In principle, we would expect to see upward pressure on storage prices until physical and synthetic storage costs converge.
13. Gold is usually in contango because of its limited industrial applications (hence, a nonexistent convenience yield). Gold briefly moved into backwardation after the adoption of the Washington Agreement on Gold in September 1999, in which European central banks agreed to restrict the sale and lending of gold.
14. Locking in low financing rates with a gold loan is synthetically equivalent to a forward sale of gold. If spot prices later rise, the mine thus cannot realize any gains from sales of its own gold at those higher prices. But mines with external credit problems for which gold loans may be most appealing might be better served by avoiding the risk of gold price fluctuations anyway. Tufano (1996) provides an insightful review of risk management practices in the gold mining industry more generally.
15. Up to the time of Hammurabi, the commodity interest rate on grain seemed to remain at around 20 percent. In the famous legal Code of Hammurabi, a maximum commodity interest rate of 33 1/3 percent per year was stipulated (Bromberg 1942).
16. Prepaid forwards can be a legitimate form of commodity-based finance (see, e.g., Culp and Kavanagh 2003). Unfortunately, Enron gave prepaid forwards a bit of a bad name. The problems with the Enron structures were not with the prepaid forwards per se but with the manner in which Enron used those structures to facilitate seemingly misleading accounting and disclosure practices.
17. See the examples in Culp (2006).
18. One of the conditions of salam is that it may not be based on a specific asset that is subject to destruction or degradation. A farmer thus cannot precontract to sell the specific crop from a specific field but can enter into an agreement to sell a specified amount and quality of comparable grain instead.
19. The use of derivatives in medieval Europe in general and by the Medici Bank in particular is discussed in great detail by De Roover (1948, 1963), from which most of the historical facts presented in this section are drawn.
20. See Culp and Heaton (2007) for a discussion of stock lending in the context of the "naked shorting" controversy.
21. See, e.g., Hafner and Wallmeier (2006).
22. Miller (1996) examines the social costs of the so-called great derivatives disasters of the 1990s. His admonitions and commentary are eerily applicable to the credit crisis, as well. See also Miller (1991).

FURTHER READING

Bakken, H. H. 1953. *Theory of Markets and Marketing*. Madison, WI: Mimir Publishers.

Culp, C. L. 2004. *Risk Transfer: Derivatives in Theory and Practice*. Hoboken, NJ: John Wiley & Sons.

Culp, C. L., and M. H. Miller. 1995. "Metallgesellschaft and the Economics of Synthetic Storage," *Journal of Applied Corporate Finance* 7, no. 4 (Winter): 62–76.

Culp, C. L., and M. H. Miller, eds. 1999. *Corporate Hedging in Theory and Practice: Lessons from Metallgesellschaft*. London: Risk Books.
Culp, C. L., and W. A. Niskanen, eds. 2003. *Corporate Aftershock: The Public Policy Lessons from the Collapse of Enron and Other Major Corporations.* Hoboken, NJ: John Wiley & Sons.
Culp, C. L., S. H. Hanke, and A. M. P. Neves. 1999. "Derivative Diagnosis," *International Economy* (May/June): 36–37, 67.
Figlewski, S. 1994. "How to Lose Money in Derivatives," *Journal of Derivatives* 2, no. 2 (Winter): 75–82.
French, K. R. 1986. "Detecting Spot Price Forecasts in Futures Prices," *Journal of Business* 59, no. 2–2 (April): S39–S54.
Kroszner, R. S. 2008. "Assessing the Potential for Instability in Financial Markets, Speech before the Risk Minds Conference, December 8.Geneva, Switzerland.

REFERENCES

Banz, R. W., and M. H. Miller. 1978. "Prices for State-Contingent Claims: Some Estimates and Applications," *Journal of Business* 51, no. 4 (October): 653–672.
Brav, A., and J. B. Heaton. 2003. "Market Indeterminacy," *Journal of Corporation Law* 28, no. 4 (Summer): 517–539.
Bromberg, B. 1942. "The Origin of Banking: Religious Finance in Babylonia," *Journal of Economic History* 2, no. 1 (May): 77–88.
Carlton, D. W. 1984. "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures," *Journal of Futures Markets* 4, no. 3 (Fall): 237–271.
Culp, C. L. 2006. *Structured Finance and Insurance*. Hoboken, NJ: John Wiley & Sons.
Culp, C. L., and J. B. Heaton. 2007. "Naked Shorting," SSRN Working Paper (August).
Culp, C. L., and B. T. Kavanagh. 2003. " Structured Commodity Finance after Enron: The Uses and Abuses of Pre-paid Forwards and Swaps," in C. L. Culp and W. A. Niskanen, eds., *Corporate Aftershock: The Public Policy Lessons from the Collapse of Enron and Other Major Corporations*. Hoboken, NJ: John Wiley & Sons.
De Roover, R. 1948. *Money, Banking and Credit in Medieval Bruges*. Cambridge, MA: Medieval Academy of America.
De Roover, R. 1963. *The Rise and Decline of the Medici Bank* (1999 ed.). Washington, D.C.: Beard Books.
Demsetz, H. 1969. "Information and Efficiency: Another Viewpoint," *Journal of Law and Economics* 12, no. 1 (April): 1–22.
Face, R. D. 1958. "Techniques of Business in the Trade between the Fairs of Champagne and the South of Europe in the Twelfth and Thirteenth Centuries," *Economic History Review* 10, no. 3: 427–438.
Fama, E. F., and K. R. French. 1987. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *Journal of Business* 60, no. 1 (January): 55–73.
Fama, E. F., and K. R. French. 1988. "Business Cycles and Behavior of Metals Prices," *Journal of Finance* 43, no. 5 (December): 1075–1093.
Froot, K. A., D. S. Scharfstein, and J. C. Stein. 1993. "Risk Management: Coordinating Corporate Investment and Financing Policies," *Journal of Finance* 48, no. 5 (December): 1629–1958.
Furbush, D. 1989. "Program Trading and Price Movement: Evidence from the October 1987 Market Crash," *Financial Management* 18, no. 3 (Autumn): 68–83.
Gammill, J. F., and T. A. Marsh. 1989. "Trading Activity and Price Behavior in the Stock and Stock Index Futures Markets in October 1987," *Journal of Economic Perspectives* 2, no. 3 (Summer): 25–44.

Gorton, G., and J. G. Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal* 35, no. 2 (March/April): 47–68.
Hafner, R., and M. Wallmeier. 2006. "Volatility as an Asset Class: European Evidence, SSRN Working Paper (January).
Harris, L. 1989. "The October 1987 S&P500 Stock-Futures Basis," *Journal of Finance* 44, no. 1 (March): 77–99.
Jackwerth, J. C., and M. Rubinstein. 1996. "Recovering Probability Distributions from Option Prices," *Journal of Finance* 51, no. 5 (December): 1611–1631.
Jastrow, M. 1911. *Aspects of Religious Belief and Practice in Babylonia and Assyria*. New York: American Academy of Religion.
Keynes, J. M. 1930. *A Treatise on Money: Volume I, The Pure Theory of Money, and Volume II, The Applied Theory of Money* (1950 ed.). London: Macmillan.
Kleidon, A. W. 1992. "Arbitrage, Nontrading, and Stale Prices: October 1987," *Journal of Business* 65, no. 4 (October): 483–507.
Kleidon, A. W., and R. E. Whaley. 1992. "One Market? Stocks, Futures, and Options during October 1987," *Journal of Finance* 47, no. 3 (July): 851–877.
Kuserk, G. J., and P. R. Locke. 1994. "The Chicago Loop Tunnel Flood: Cash Pricing and Activity," *Review of Futures Markets* 13, no. 1: 115–146.
Merton, R. C. 1992. "Financial Innovation and Economic Performance," *Journal of Applied Corporate Finance* (Winter): 12–22.
Miller, M. H. 1991. "Leverage," *Journal of Finance* 46, no. 2 (June): 479–488.
Miller, M. H. 1996. "The Social Costs of Some Recent Derivatives Disasters," *Pacific Basin Finance Journal* 4 (July): 113–127.
Mulherin, J. H., J. M. Netter, and J. A. Overdahl. 1991. "Prices Are Property: The Organization of Financial Exchanges from a Transaction Cost Perspective," *Journal of Law and Economics* 34, no. 2-2 (October): 591–644.
Roll, R. 1984. "Orange Juice and Weather," *American Economic Review* 74, no. 5 (December): 861–880.
Smith, F. L. Jr. 2003. "Cowboys versus Cattle Thieves: The Role of Innovative Institutions in Managing Risks along the Frontier," in C. L. Culp and W. A. Niskanen, eds., *Corporate Aftershock: The Public Policy Lessons from the Collapse of Enron and Other Major Corporations*. Hoboken, NJ: John Wiley & Sons.
Sraffa, P. 1932. "Dr. Hayek on Money and Capital," *Economic Journal* 42, no. 165 (March): 42–53.
Stoll, H. R., and R. E. Whaley. 1990. "The Dynamics of Stock Index and Stock Index Futures Returns," *Journal of Financial and Quantitative Analysis* 25, no. 4 (December): 441–468.
Swan, E. J. 2000. *Building the Global Market: A 4000 Year History of Derivatives*. London: Kluwer Law.
Telser, L. 1958. "Futures Trading and the Storage of Cotton and Wheat," *Journal of Political Economy* 66, no. 3 (June): 233–255.
Tufano, P. 1996. "Who Manages Risk? An Empirical Examination of Risk Management Practices in the Gold Mining Industry," *Journal of Finance* 51, no. 4 (September): 1097–1137.
Varchaver, N., and K. Benner. 2008. "The $55 Trillion Question," *Fortune*, September 30, 2008.
Williams, J. 1986. *The Economic Function of Futures Markets*. Cambridge, MA: Cambridge University Press.
Working, H. 1948. "Theory of the Inverse Carrying Charge in Futures Markets," *Journal of Farm Economics* 30, no. 1 (February): 1–28.
Working, H. 1949. "The Theory of Price of Storage," *American Economic Review* 39, no. 6 (December): 1254–1262.

ABOUT THE AUTHOR

Christopher L. Culp is senior advisor with Compass Lexecon, an adjunct professor of finance at the University of Chicago Booth School of Business, an adjunct professor at Universität Bern in the Institut für Finanzmanagement, and director of Risk Management Consulting Services, Inc. He teaches graduate degree courses on structured finance, insurance, and derivatives, and provides advisory consulting services and testimonial expertise in these same areas. Dr. Culp is the author of four books—*Structured Finance & Insurance*; *Risk Transfer: Derivatives in Theory & Practice*; *The ART of Risk Management*; and *The Risk Management Process*—and the coeditor of two books—*Corporate Aftershock: The Public Policy Consequences of the Failure of Enron and Other Major Corporations* (with William Niskanen) and *Corporate Hedging in Theory & Practice* (with Merton Miller). He holds a PhD in finance from the University of Chicago Booth School of Business.

PART II

Types of Financial Derivatives

Derivatives markets originated in the United States in the middle of the nineteenth century with contracts based on agricultural products, and the range of underlying instruments was quite limited until the middle of the twentieth century, when metallurgical derivatives were introduced. The first specifically *financial* derivative contract on an organized U.S. exchange was introduced only in 1973, with foreign currency futures contracts that began trading on the Chicago Mercantile Exchange.

The chapters in Part Two introduce the current wide variety of financial derivatives. This part begins with a discussion of agricultural and metallurgical derivatives, which are not true financial derivatives in the strictest sense, as the underlying goods for these derivatives are physical, not financial, products. Nonetheless, the understanding of derivatives was first developed with these contracts, and they share common pricing principles with financial derivatives in particular. Joan C. Junkus provides the analysis of agricultural and metallurgical derivatives in Chapters 5 and 6, "Agricultural and Metallurgical Derivatives: Pricing," and "Agricultural and Metallurgical Derivatives: Speculation and Hedging." Junkus surveys the complete range of issues that arise with respect to these fundamental derivatives.

Derivatives based on equities are a fundamental kind of specifically financial derivative, with the principal underlying good being a stock index, such as the Standard & Poor's 500 stock market index. Jeffrey H. Harris and L. Mick Swartz survey these markets in Chapter 7, "Equity Derivatives." Because of their intimate connection with the stock market, equity derivatives are widely used by speculators and hedgers. Portfolio managers use equity derivatives to shape the risk and return characteristics of their portfolios to get exactly the kind of anticipated distribution of payoffs and risks that they desire. Equity derivatives have become extremely popular around the world, with these instruments playing a prominent role in virtually all national derivatives markets.

In Chapter 8, "Foreign Exchange Derivatives," Robert W. Kolb notes the enormous size of the foreign exchange markets and goes on to explain the basic principles that govern the pricing of foreign exchange derivatives, including the purchasing power parity and interest rate parity theorems. In the foreign exchange market, the over-the-counter market dwarfs exchange-traded derivatives, and this is true for forwards, options, and swaps, all of which Kolb discusses.

In recent years, the prices of petroleum products have exhibited violent swings, emphasizing the importance of Chapter 9 by Craig Pirrong, "Energy Derivatives." Introduced only in the 1970s, these instruments have grown in importance both on organized exchanges as well as in the over-the-counter market. In terms of market size, petroleum products dominate other derivatives, but these other derivatives include propane, natural gas, and electricity. Closely related is a market on sulfur dioxide emissions, a by-product of electricity generation. Pirrong evaluates these important markets with respect to the types of products traded, the principles of energy derivatives pricing, and the relationships among prices for various energy derivatives.

Another recently developed market of rapidly growing importance is the market for derivatives tied to interest rates, which Ian Lang evaluates in Chapter 10, "Interest Rate Derivatives." In most instances, the actual good that underlies an interest rate derivative is a debt instrument, such as a money market deposit or a Treasury bond. As Lang notes, these markets are now dominated by the over-the-counter market, with the full range of derivatives (futures, forwards, options, swaps, etc.) being represented. Lang goes on to show how these instruments can be used, either singly or in combination, to take and avoid interest rate risk.

As financial markets have matured, the variety of products available has gone far beyond the plain vanilla instruments of forwards, futures, options, and swaps, to embrace a class of exotic options, which Robert W. Kolb surveys in Chapter 11, "Exotic Options." The exotic category embraces a tremendous range of derivatives ranging in complexity from those that are quite simple to those that are extremely complex. Exotic options are traded almost exclusively in the over-the-counter market, with quite robust markets being available for certain kinds of exotics. As Kolb shows, the pricing principles that apply to plain vanilla options can be extended to the pricing of exotic options with great success.

Justin Wolfers and Eric Zitzewitz analyze one particularly important class of derivatives in Chapter 12, "Event Derivatives." An event derivative is a contract that pays off if and only if a well-defined event occurs. These kinds of instruments have achieved a particular prominence in terms of election politics with contracts trading on events such as "Obama wins the presidency." Wolfers and Zitzewitz analyze these instruments and the markets in which they trade. As the authors show, the price of an event derivative may be interpreted as reflecting the market's assessment of the probability that a particular event will occur.

In 2006, few people outside of financial markets had ever heard of credit default swaps, the topic of Chapter 13 by Steven Todd. To some extent, awareness of these instruments has penetrated ordinary discourse in the wake of recent financial difficulties. In essence, a credit default swap is a contract that pays off if a credit event, such as a default, occurs. Todd explains that credit default swaps are the building blocks for all more complex credit derivatives. He also goes on to discuss the pricing of these instruments and the most important credit default swap indexes.

As a companion to his chapter on credit default swaps, Steven Todd's Chapter 14, "Structured Credit Products," extends the analysis to more complex credit derivatives such as asset-backed securities, collateralized debt obligations, and commercial mortgage-backed securities, all of which have played important roles in the financial crisis. As Todd notes, "One feature common to all structured credit products is the use of financial engineering techniques to create securities that provide a range of risk-return profiles for different investors." These markets have grown extremely rapidly, attesting to the uses that they serve, although their future seems less secure in the light of recent financial difficulties.

While financial derivatives are almost always controversial, perhaps the most controversial of all financial derivatives are described in Chapter 15, "Executive Stock Options," by Robert W. Kolb. Given the widespread controversy over executive pay, executive stock options are at the center of the debate. They are the principal vehicle for conveying wealth to executives, dwarfing other components of executive compensation such as salary, bonus, retirement packages, and perquisites. After surveying the components of executive pay, Kolb considers the rationales for executive stock options and discusses some of the pricing principles for them. They turn out to be quite complex, and executive stock options are difficult to price for a variety of reasons.

To conclude Part Two, Steve Swidler looks to the future of derivatives in Chapter 16, "Emerging Derivatives Instruments." Swidler focuses on the potential for two classes of derivatives, economic derivatives and real estate derivatives. Economic derivatives include contracts on such phenomena as inflation and other macroeconomic indexes. As Swidler notes, some of these contracts have been tried with limited success, so he evaluates the conditions that would be likely to lead to their market acceptance. He applies a similar analysis to real estate derivatives, some of which have been marketed without success. He notes that success of any such contract requires hedging effectiveness, for those who wish to use these markets to lay off risk. But a successful contract must also appeal to speculators to induce them to provide the liquidity that hedgers demand.

CHAPTER 5

Agricultural and Metallurgical Derivatives

Pricing

JOAN C. JUNKUS
Associate Professor of Finance, DePaul University, Chicago

INTRODUCTION

This chapter explores the major elements of futures pricing as they relate to commodities. We begin with the seasonal price behavior of commodities. Next, we review two major pricing models. The theory of storage uses supply and demand behavior in the underlying cash commodity market to price futures. The theory of normal backwardation looks to hedging behavior to explain futures prices. Finally, we summarize the empirical evidence for these two theories.

COMMODITIES

Commodities are held or stored in inventory for use as inputs in a production process. Unlike financial assets, they are valued based on the their future expected spot prices rather than future expected cash flows. Mismatches between production of a commodity by harvest or mining and its consumption in a production process can lead to seasonality in commodity prices. Unlike the case with financials, such seasonals with commodities are not necessarily market inefficiencies.

The value of a futures contract on a commodity is based primarily on expected future spot prices and the storage costs of holding the particular commodity. Thus, commodity prices and their derivatives are determined by supply and demand behavior in a particular market. Understanding the pricing of a commodity can require detailed knowledge of supply and demand factors in the cash market. Because derivatives pricing depends on how spot prices and anticipated spot prices are determined, seasonality in commodity spot and futures pricing is discussed first in this chapter, and then the two major pricing theories for commodities are outlined along with a review of the empirical evidence.

SEASONALITY IN SPOT AND FUTURES PRICES

Commodities prices often exhibit patterns related to the seasonality of production or consumption. Many agricultural commodities are harvested once a year, so supply is fixed for that particular crop year while consumption occurs continuously. A typical seasonal price pattern is depicted in Exhibit 5.1. Prices serve as signals to the market as to how to allocate this crop inventory efficiently over the crop year. At harvest (H), the spot price (S_H) is at its lowest as the supply of the commodity peaks. After harvest, spot prices during the rest of the year are expected to be higher, reflecting the current spot price plus the cost of storage to that point in time. At the midpoint of the year, the spot is expected to be S_{MP}, which is also equal to S_H plus the cost of storage from harvest up to the year's midpoint. At the next harvest, new supply becomes available, and prices drop to reflect the increased inventory. Thus, the underlying seasonal price pattern is roughly saw-toothed, with a trough at harvest when supply is highest and ending in a peak at the end of the crop year when supply is at its lowest just before the next harvest.

The seasonal pattern, though predictable, is not necessarily evidence of market inefficiency. Seasonal price patterns are not as sharply defined as the discussion implies because crops vary in how successfully they can be stored. More important, seasonality will be obscured by factors that affect the spot price (and future expected spot prices) as the crop year unfolds. New information will shift the expected cash price pattern in Exhibit 5.1 up or down, and such changes in prices can overwhelm the seasonal pattern as the crop year develops. Inventory may also be carried from previous crop years, and this larger supply lowers expected spot prices and dampens the seasonal. Similarly, if more supply can be brought into the market during the crop year (from, e.g., a different climate zone), this will also disrupt the seasonal pattern.

The futures price for delivery at time t in the crop year will be equal to the expected spot price, which in turn is equal to the current spot price plus expected

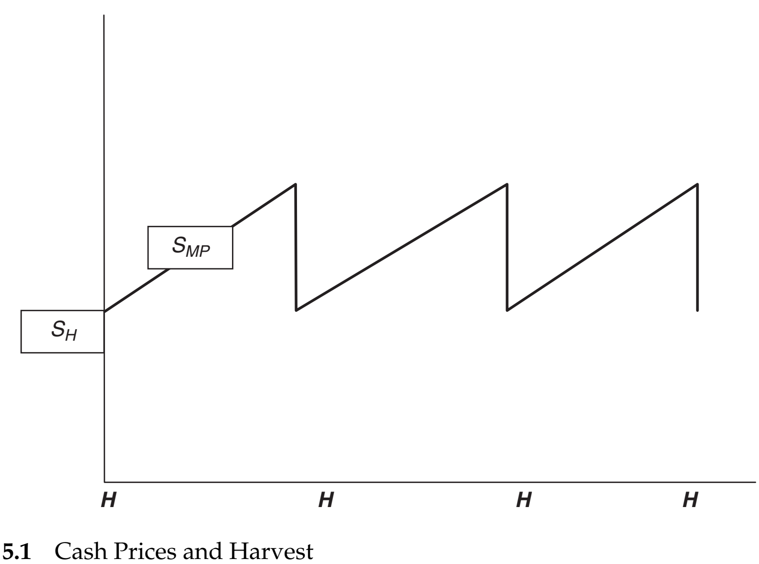


Exhibit 5.1 Cash Prices and Harvest

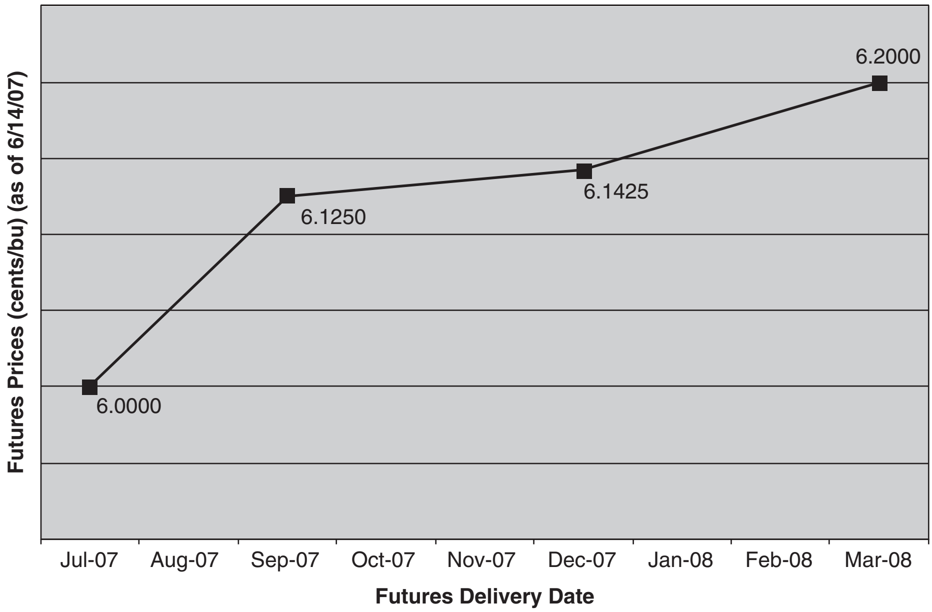


Exhibit 5.2 Contango in Wheat Futures Prices (Kansas City Board of Trade Winter Wheat)

storage costs to period t. The futures curve, or futures strip, is the set of futures prices for different maturities, and it will be upward sloping (called contango) in this case. Exhibit 5.2 shows a strip of wheat futures in contango. The futures price for deferred delivery will be greater than the futures price for nearby delivery, reflecting the market's estimate of the expected change in cash prices over that time. The difference between futures prices also reflects the return to storage; that is, what the market is willing to pay the holder of inventory to store the asset over the crop year. In this way, futures prices (expected cash prices) will allocate the crop inventory over the crop year. As long as the futures price/expected cash price covers the current cash price plus storage costs, the crop will be stored.

The basis of a futures contract is defined as the current spot price minus the futures price ($SP - FP$). The basis reflects expected storage costs between today and the futures' maturity. As the futures contract approaches maturity, storage costs approach zero, the futures price will converge to the cash price, and the basis will approach zero as well. The basis will thus also have a distinct seasonal component, and many producers use basis seasonality to make hedging decisions. However, because new information will alter the expected future cash price, the basis also fluctuates in response to information. Over time, changes in the basis will reflect both an expected change (storage costs), and an unexpected change (new information).

FUTURES PRICING

Two models have been proposed to explain the pricing of commodity futures. The theory of storage, discussed first, places major emphasis on the link between cash prices and forward prices, with forward prices reflecting physical storage costs and an imputed convenience yield arising from a mismatch between anticipated supply and consumption demand. The alternative theory of forward pricing, normal backwardation, places greater emphasis on risk management and the reward for risk.

Theory of Storage

The key differences between pricing futures on financial assets and commodities are the cost associated with carrying a commodity through time, and the cost of borrowing a physical asset in an arbitrage. For a commodity that is easily stored, such as gold, and with plentiful supply relative to demand, the futures contract will be priced at full carry over the spot price, and the differences between spot and futures price, or the futures prices for different maturities, reflect financing, physical storage, and insurance costs for that time period. For a full-carry commodity, the basic no-arbitrage framework is the same used to price forwards on financial assets. If the forward price (FP) is greater than the spot price and carrying costs $[SP + cSP]$, then arbitrageurs can enter into a cash-and-carry arbitrage. They can obtain an arbitrage profit by selling the overpriced forward contract at FP, buying the spot commodity, financing the position, and paying the storage costs. At maturity, the arbitrage profit would be $FP - [SP + cSP]$. Thus, $[SP + cSP]$ is the upper bound for the forward price. The difference between forward prices for different maturity contracts can be expressed as a percentage $[(FP_2 - FP_1)/FP_1]$, and this implied repo rate will be equal to the implied forward financing rate for that time period plus physical storage costs.

In the reverse cash and carry, if the FP is less than $[SP + cSP]$, then arbitrageurs should buy the forward, sell short the gold, and invest the spot price cash flow at the financing rate until the forward contract matures. With a commodity, however, there is a charge to the arbitrageurs to borrow the physical asset, called the lease rate. The commodity owner demands a lease rate because a commodity does not necessarily increase in value as a financial asset is expected to do, and the commodity owner will want compensation for lending a real asset. While an investor in a financial asset expects to receive a positive, risk-adjusted rate of return, whether through dividends or capital gains, even if the asset is borrowed to enable a short sale, a commodity owner does not. Thus, the lower bound for a commodity futures price incorporates a lease rate for lending the physical asset.

In a full-carry commodity such as gold, the futures price will center around the equilibrium forward price in a band that reflects transaction costs, the ease of selling the commodity short (reflected in the lease rate), and other imperfections in the arbitrage transactions.

Convenience Yield and the Theory of Storage

Gold is generally freely available for borrowing, the lease rate is known and published, and arbitrage is relatively easy to do. Many commodities, however, are difficult and costly to store. More important, the primary purpose of holding most commodities in inventory is to use the commodity as an input in a production process. It is the need to have physical possession of the commodity that makes the forward pricing of commodities distinctively different from financials.

For these commodities, a cash-and-carry arbitrage presents no special problems, for if the futures price is greater than the spot price plus carry costs $[SP + cSP]$, the arbitrageur sells the forward, buys the spot, incurs the financing and storage costs, and makes an arbitrage profit of $FP - [SP + cSP]$.

However, a reverse cash and carry depends critically on the ability to borrow the commodity from someone who is storing it. For commodities such as gold, borrowing is not difficult, and the lease rate charged to the arbitrageur to borrow the asset will be approximately equal to the storage costs the owner saves by lending it to the arbitrageur over the life of the contract. The lower bound on the futures price will thus reflect spot price, financing costs, and the lease rate/storage costs saved by the commodity asset owner.

But some commodities cannot be easily stored or may have inadequate supplies relative to the quantity needed in a production process. In this case, the asset owner may not be willing to part with the physical commodity just to save storage costs. The owner will be concerned with stock-outs, the disruption of production, and the cost of shutting down a production process until input is found, if it can be purchased at all. This reluctance to lend out inventory, the value the asset owner assigns to actual physical possession, is referred to as the convenience yield. In cases where supply is expected to be low relative to expected demand or where storage is difficult, the value of physical possession, or convenience yield, will be large and cause the forward price to be less than full carry. Convenience yield is unobservable but its existence is implied when futures prices are at less than full carry: If commodity holders continue to store inventory when there is a negative return to storage (the futures price is less than the spot price and storage costs), it is argued that a convenience yield must compensate the owner for this negative return.

Convenience yield is the benefit or return to the owner of inventory from the flow of services expected from holding the commodity. As such, convenience yield can accrue only to the commodity holder: Someone long the *futures* contract (such as the arbitrageur in the reverse cash and carry) cannot capture this part of the commodity value. The lower bound on the futures price thus reflects the current spot price, plus financing and storage costs, *less* the convenience yield. As a benefit from holding the commodity, the convenience yield may also be considered a negative cost.

Since the size of the convenience yield is the key to commodity pricing, it is useful to ask under what circumstances a commodity will be stored. In classical economics, where supply and demand equilibrate through price, a production equilibrium (price has adjusted so that suppliers produce enough to satisfy consumer demand at that price) does not involve storage of the commodity. Storage arises only if there is a temporal mismatch between the production/supply of the commodity and its consumption/demand. Harvested commodities will be stored because periodic harvest must be matched with constant consumption. This is a subtle point, but it emphasizes the fact that determining the equilibrium futures price for a commodity and hence the size of the convenience yield involves analyzing the decision whether to store the commodity. The behavior of the convenience yield and the futures price will be strongly determined by a commodity's specific expected demand and supply. Further, since storage also connects harvest years (old crop versus new crop), the size of successive harvests and excess inventory

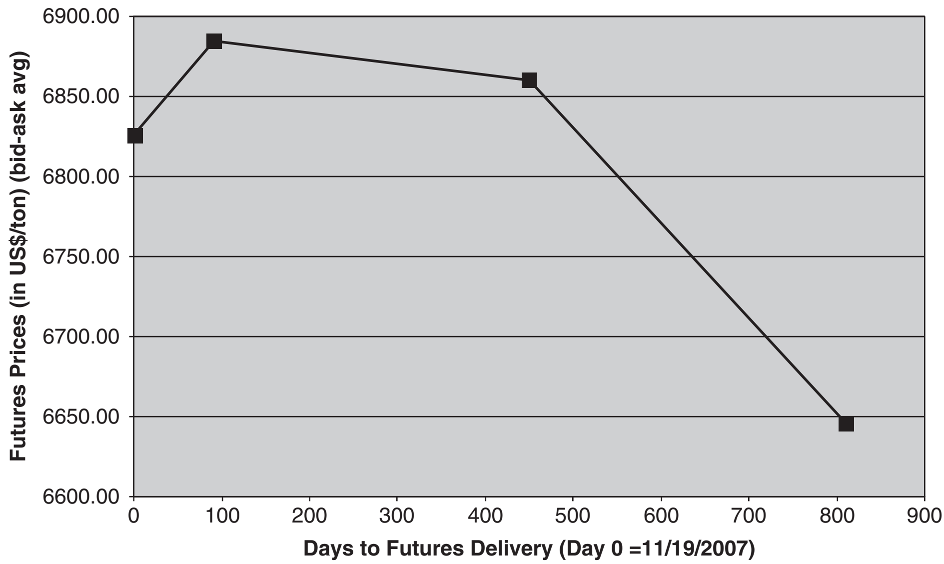


Exhibit 5.3 Backwardation in Copper Prices (London Metal Exchange)

expected to be stored between harvest years will also have an impact on convenience yield and the futures price.

The basic theory of storage implies that the size of the convenience yield and its behavior through time will be related to several factors. It is dependent on the size of storage costs in relation to commodity value. Some commodities, such as copper, are costly to store because of their bulk in relation to their value. Although storage is possible, it is usually minimal in relation to usage. These commodities frequently trade at less than full carry or even at backwardation. Backwardation (also called inverse carry) occurs if futures prices for distant delivery are less than spot prices or futures prices for nearby delivery. See Exhibit 5.3 for an example of backwardation in copper futures prices.

Other commodities are difficult to store due to spoilage (orange juice) or aging (livestock, live hogs). At the extreme, the storage costs of these commodities, and the convenience yield of possession, can be considered infinite. For such nonstorable commodity assets, the forward price is simply equal to the expected spot price at maturity. For these commodities, the futures curve (and basis) exhibits little coherent pattern. Without the ability to store the commodity, there is little linkage between spot prices and future delivery prices. For nonstorables, the futures price is equal to the expected spot price, and contango and backwardation are equally likely.

Convenience yield, and departure from full carry prices, also depends on the scarcity of the commodity and should increase as inventory levels become low in relation to expected demand. Thus, convenience yield should be negatively correlated with inventory levels at a nonlinear, decreasing rate. This fact also implies that convenience yield should vary with the business cycle. Inventory responsiveness, or supply elasticity, should be at its lowest when all productive inputs are relatively scarce as the business cycle peaks, so convenience yield should be high at this time. Last, since spot prices adjust to demand shocks, convenience yield also should be positively correlated to spot prices: Higher spot prices signal high demand relative to fixed supply, which should lead to increased convenience yield.

A more recent version of the theory of storage models considers the convenience yield as an embedded timing option for the commodity holder and posits a switching behavior in spot prices. The commodity owner can decide to store the commodity in inventory, in which case it is priced as an ordinary asset, and the forward price reflects current prices, storage costs, and thus expected prices. If it is optimal, however, to consume the commodity (or sell it into the spot market), then the commodity is priced as a consumption good, a convenience yield is embedded in the forward price, and the linkage among current prices, storage costs, and expected prices is broken (Routledge, Seppi, and Spatt 2000).

Since the convenience yield reflects the value of the inventory holder's option to sell the commodity (a put option), then the holder's position (put option plus long commodity) and the size of the convenience yield can be valued as a call option (Milonas and Thomadakis 1997). As a call option, convenience yield should be positively related to spot price volatility, the exercise price (the futures price available to the commodity holder if he or she sells), and the marginal cost of production (which affects the supply response to demand shocks). However, the convenience yield should be negatively related to serial correlation in spot prices: If a relatively high spot price is not expected to continue beyond the future's maturity—indicating low spot price serial correlation—then holding the commodity rather than taking a futures position yields a high convenience yield (Heinkel, Howe, and Hughes 1990).

There is some dissatisfaction with using an implied and unobservable convenience yield to explain an apparent negative return to storage, and several alternative explanations have been proposed to account for less than full carry in futures prices. Convenience yield may be an artifact of aggregated data and the result of mismeasurement (Wright and Williams, 1989). Under this approach, the appearance of a negative return to storage can arise through aggregating localized prices and the behavior of localized warehouses with positive transportation costs or quality differences. Other explanations for an apparent negative return to storage include transaction costs (high transactions costs, particularly during periods of low inventory, may by themselves supply an incentive to store) (Chavas, Despins, and Fortenbery 2000) and asymmetric information between commodity holders (hedgers) and speculators (Frechette 2001).

Empirical Tests of the Theory of Storage

Determining the size and behavior of the convenience yield has grown in importance as commodity forward prices increasingly are used as inputs for risk management programs and for real options models used to value capital projects involving exhaustible resources. Correlation between convenience yield and spot price levels, for instance, implies a mean reversion process in spot prices and correspondingly lower real option values. Similarly, capturing the dynamics of commodity price behavior is vital to calculate risk measures such as value at risk (Casassus and Collin-Dufresne 2005).

Convenience yield and its dynamics have been difficult to test empirically for several reasons. Since convenience yield must be inferred from futures prices and estimated storage costs, it is sensitive to how storage costs are estimated. Further, measuring relative scarcity requires information on inventories, which is generally incomplete and reported with variable lags. With these limitations, most studies have found a negative relation between convenience yield and inventory levels or measures of scarcity (Dincerler, Khokher, and Simin 2005; Fama and French 1987).

The tests relating convenience yield to option characteristics have been mixed. There is some evidence that convenience yield is related to the exercise price (Milonas and Thomadakis 1997), but other results cast doubt on the option price model (Sorensen 2002). There also appears to be a relatively complex relationship between convenience yield and spot prices (positively correlated to convenience yield, although time-varying), demand shocks, and relative scarcity (negatively correlated to inventory).

Theory of Normal Backwardation

According to the theory of normal backwardation, speculators in commodity futures will receive a positive return as compensation for the price risk transferred to them by the hedger. Keynes's (1930) original theory assumed that the hedger, as a producer, would sell futures. The risk premium would be paid to the long speculator by setting the futures price below the expected future spot price: The futures prices would, on average, rise through time, resulting in a positive rate of return for the long speculator.

Note that this use of the term *backwardation* differs from that used in describing the basis or futures curve. *Backwardation* occurs when the current spot price is greater than a futures price. *Normal backwardation*, however, describes the relationship between the forward price and the expected spot price: The forward price is set below the expected spot price, and this persistent downward bias in the forward price rewards the long speculator with positive returns.

A modification of the normal backwardation model argues that not all markets will return a risk premium to the speculator (Telser 1958). Competition between speculators normally will drive the risk premium to zero, so only markets in which there are too few speculators willing to take on hedger risk will pay a risk premium to a long position. Such a thin market would be characterized by low liquidity, with few transactions, a low volume of trade in the commodity relative to its production, and consequently large price fluctuations.

Keynes's original model has been modified to recognize that not all hedging interest will be short. Hedgers can be producers *or* consumers, and speculators will garner positive returns if they take a position opposite to the majority of hedgers in the market. If hedgers are net short, then a long speculative position should result in a positive return, as in the original backwardation model. If hedgers are net long in a particular market, then speculators are paid to go short, and a short position will reap positive returns. Thus, to reap consistently positive returns, speculators must be able to tell where hedging net interest will be in a particular market.

It is important to recognize that the risk premium on commodity futures contracts can be linked to convenience yield. We have seen that if futures prices are at less than full carry because of inadequate inventories, then convenience yield will be relatively large; this implies increased uncertainty and a demand for risk transference. If speculators are to be rewarded for providing a risk transfer function, then periods of high convenience yield (when futures are driven below the current spot price—the traditional definition of backwardation) may also be periods in which long futures positions exhibit positive returns. The size of the risk premium will depend on inventory levels, which drive the convenience yield, and the relative risk sensitivities of inventory holders (hedgers) and investors (speculators).

This relationship between futures returns and risk premiums exhibits momentum, since inventory shocks and deviations from average normal inventory levels will require considerable time to adjust where commodities are produced only seasonally. In other words, risk premiums can be expected to persist for some time, and price behavior (the adjustment of spot and futures prices to inventory shocks) can be a predictor of these prolonged excess returns.

Although commodities are technically real rather than financial assets, asset pricing models have also been applied to commodity futures. Under the capital asset pricing model, for instance, a fully collateralized futures contract should exhibit returns commensurate to its systematic risk. Hirschleifer (1988) has proposed a model that combines systematic risk with net hedging. In this model, if there are market imperfections that impede trading, then a commodity's risk premium will consist of systematic risk plus or minus a residual factor that is related to speculator risk aversion and the degree of market imperfection. This residual factor is added to systematic risk if hedgers are net short, and a positive risk premium will be associated with a net short hedging position.

Empirical Tests of Normal Backwardation

Is there a risk premium in futures markets? Most empirical studies find that commodities exhibit low, or even negative, correlation with market portfolio proxies (Gorton and Rouwenhorst 2006; Jensen, Johnson, and Mercer 2000). While empirical studies have tended to find that commodity futures do not have significant systematic risk (Dusak 1973), there continues to be debate about the proper definition of the market portfolio. In addition, risk premiums have been found to change with monetary regime and the business cycle (Bessembinder and Chan 1992; Bjornson and Carter 1997).

Risk premiums in futures are also dependent on net hedging. Since net hedging positions are unobservable, however, testing this version of the theory has relied on using CFTC-supplied large hedger positions as a proxy. Consistent with the net hedging model, commodity returns have been found to be positively related to the net position, long or short, of hedgers in a particular market (DeRoon, Nijman, and Veld 2000; Kolb, 1992).

CONCLUSION

Unlike financial assets, agricultural commodities exhibit seasonality in prices, and this seasonality has a profound effect on the price behavior of agricultural futures. One of the two models of futures pricing, the theory of storage, explains futures prices by the need to allocate seasonal production across the year or to carry precautionary inventory. Normal backwardation, however, emphasizes the risk management function. This chapter reviewed both pricing theories and the empirical evidence. See also Chapters 9 and 25.

REFERENCES

Bessembinder, H., and K. Chan. 1992. "Time-Varying Risk Premia and Forecastable Returns in Futures Markets," *Journal of Financial Economics* 32: 169–193.

Bjornson, B., and C. A. Carter. 1997. "New Evidence on Agricultural Commodity Return Performance under Time-Varying Risk," *American Journal of Agricultural Economics* 79: 918–930.

Casassus, J., and P. Collin-Dufresne. 2005. "Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates," *Journal of Finance* 55: 2283–2331.

Chavas, J. P., P. M. Despins, and T. R. Fortenbery. 2000. "Inventory Dynamics under Transaction Costs," *American Journal of Agricultural Economics* 82: 260–273.

DeRoon, F. A., T. E. Nijman, and C. Veld. 2000. "Hedging Pressure Effects in Futures Markets," *Journal of Finance* 55: 1437–1456.

Dincerler, C., Z. Khokher, and T. Simin. 2005. "An Empirical Analysis of Commodity Convenience Yields," (June); available at: http://ssrn.com/abstract=748884.

Dusak, K. 1973. "Futures Trading and Investment Returns: An Investigation of Commodity Market Risk Premiums," *Journal of Political Economy* 81: 1387–1406.

Fama, E. F., and K. R. French. 1987. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *Journal of Business* 60: 55–73.

Frechette, D. L. 2001. "Aggregation and the Nature of Price Expectations," *American Journal of Agricultural Economics* 83: 52–63.

Gorton, G., and K. G. Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal* (March-April): 47–68.

Heinkel, R., M. E. Howe, and J. S. Hughes. 1990. "Commodity Convenience as an Option Profit," *Journal of Futures Markets* 10: 519–533.

Hirschleifer, D. 1988. "Residual Risk, Trading Costs and Commodity Futures Risk Premia," *Review of Financial Studies* 1: 173–193.

Jensen, G. R., R. R. Johnson, and J. M. Mercer. 2000. "Efficient Use of Commodity Futures in Diversified Portfolios," *Journal of Futures Markets* 20: 489–506.

Keynes, J. M. 1930. *A Treatise on Money*. New York: Harcourt, Brace.

Kolb, R. W. 1992. "Is Normal Backwardation Normal?" *Journal of Futures Markets* 12: 75–91.

Milonas, N. T., and S. B. Thomadakis. 1997. "Convenience Yields as Call Options: An Empirical Analysis," *Journal of Futures Markets* 17: 1–15.

Routledge, B., D. Seppi, and C. Spatt. 2000. "Equilibrium Forward Curves for Commodities," *Journal of Finance* 55: 1297–1338.

Sorensen, C. 2002. "Modeling Seasonality in Agricultural Commodity Futures," *Journal of Futures Markets* 22: 393–426.

Telser, L. G. 1958. "Futures Trading and the Storage of Cotton and Wheat," *Journal of Political Economy* 66: 233–225.

Wright, B. D., and J. C. Williams. 1989. "A Theory of Negative Prices for Storage," *Journal of Futures Markets* 9: 1–13.

SUGGESTED FURTHER READING

Kamara, A. 1982. "Issues in Futures Markets: A Survey," *Journal of Futures Markets* 2: 261–294.

Pindyck, R. S. 2001. "The Dynamics of Commodity Spot and Futures Markets: A Primer," *Energy Journal* 22: 1–29.

ABOUT THE AUTHOR

Joan C. Junkus is an associate professor of finance at DePaul University in Chicago. She received both her MBA and PhD from the University of Illinois in Urbana. The author of numerous articles on futures pricing and hedging in such journals as the *Journal of Futures Markets* and *Derivatives Use, Trading & Regulation*, she serves on the board of *Managerial Finance*. Dr. Junkus is also the coauthor of the textbook *The Theory and Practice of Futures Markets*. Currently pursuing her primary interest in the history of financial markets and the development of alternative investments, she has recently published in *Financial History*. Dr. Junkus has served as associate dean of the College of Commerce at DePaul as well as chair of the Finance Department. Earlier, Dr. Junkus was on faculty in the Graduate School of Management at the University of California-Irvine and worked for Salomon Brothers in Chicago.

CHAPTER 6

Agricultural and Metallurgical Derivatives

Speculation and Hedging

JOAN C. JUNKUS
Associate Professor of Finance, DePaul University, Chicago

INTRODUCTION

This chapter explores the ways in which commodities are used in speculating and in risk management. First, we cover the basics of commodity investment, including the construction of commodity price indexes, the various sources of investment return for a passive investment strategy, and active strategies. We next review the empirical evidence concerning commodities' performance as an asset class, a diversifier, and an inflation hedge. Two approaches to hedging are outlined. In one approach, the hedger uses the basis to time commodity purchases or sales, while the second entails calculating an optimal hedge ratio. Last we discuss commodity spreads.

COMMODITIES

Commodities can be considered the largest nonfinancial market in the world. Commodities include a wide array of goods, each with its particular demand and supply fundamentals. Metals commodities include precious metals (gold and silver), which are both financial assets and industrial inputs, and industrial metals, such as copper, the platinum group, nickel, and aluminum. "Soft" commodities are a wide-ranging group, including beverage commodities (cocoa, tea, orange juice), sugar, fibers (cotton, jute), oil seeds (palm oil, canola, soybeans), and grains (rice, sorghum, corn, wheat). Other commodities are meat and livestock (broilers, hogs, cattle) and wood products (lumber, plywood, hardwood pulp).

As with any economic good, the demand for agricultural commodities and metals is determined by consumer attributes (including tastes, demographics, and disposable income), the price of substitute goods, and the market for products that use the commodity as an input. Supply depends on the yield possible under current technology, the cost of production, carryover of stocks, and exports or imports of the commodity. In some cases, government programs may be designed to support price or subsidize production or consumption. While changes in demand can occur fairly rapidly, changes in supply depend on relatively longer gestation cycles or exploration and production processes.

For agricultural commodities, weather can have a profound impact on prices. For metals, mining and refining technology is a key determinant of price. While agricultural commodities have a production cycle tied to natural growing or gestation periods, a metal's supply cycle is determined primarily by the long development times before discoveries and new capacity can come on line. Commodities prices, particularly for metals, also seem to follow a long-term, 20- to 30-year cycle, as demand triggers high prices, infrastructure investment, eventual overcapacity, and resulting declining prices. Such price behavior provides two supporting arguments for commodities investment: the existence of price trends (both long and short term) and commodities' correlation to price cycles and inflation.

DERIVATIVES

Futures contracts on agricultural commodities are among the oldest derivatives, and futures on commodities have been traded in the United States on the Chicago Board of Trade (now the CME Group [CMEG]) since the 1840s. Metals were first traded at the Commodities Exchange in the 1920s, while options on commodity futures were introduced in 1981. Contract specifications often reflect characteristics of the underlying spot commodity. For instance, the size of many U.S. contracts is determined by the amount of the commodity that fits in a standard U.S. railcar, and contracts are still called cars. The sequence of delivery months may be determined by harvest (supply) or consumption (demand). For example, CMEG corn futures' maturities are December, March, May, July, and September and correspond to significant months in the corn growing cycle.

As with all derivatives, changes in regulation and competition have resulted in increased commodity trading volume, the introduction of many new contracts (e.g., base metals at the London Metals Exchange, spices at the MultiCommodity Exchange in India), exchange consolidation (the Chicago Mercantile Exchange and Chicago Board of Trade merger), and new exchanges (Dubai Commodities Exchange in 2005, Chicago Climate Exchange in 2003). The declining cost of electronic platforms and increased information flow have attracted new participants, both institutional investors such as hedge funds and pension funds, and private investors, both private wealth and retail.

COMMODITY INVESTMENT STRATEGIES

Commodity Indexes

Commodity indexes are used to track commodity prices and to represent a portfolio of commodity positions. Since commodities are extremely heterogeneous, the behavior of a particular index, and comparisons of index performance, can be very sensitive to how a given index is constructed. Indexes vary by the selection and weighting of constituent components, and how components are rebalanced.

The Standard & Poor's Goldman Sachs Commodity Index (SP-GSCI) and the Dow Jones-AIG Commodity Index (DJ-AIGCI) are the most widely used in structuring tradable commodity index products, and futures are traded on these indexes at the CMEG. Other indexes include the Reuters/Jeffries Commodity Research Bureau (CRB) and the Deutsche Bank Liquid Commodity Index (DBLCI).

All the indexes just mentioned include a range of commodity sectors, but specific commodities and their weights differ. Component selection involves a trade-off between the index's ability to represent commodities as an asset class and the ease with which the index can be replicated. For instance, the CRB gives a broad picture of overall commodity price movements while the DJ-AIG selects components based on the liquidity of the futures contract. The SP-GSCI contains the largest number of commodities (25), while the DJ-AIG constrains the weighting that any individual sector can have in the index. Many indexes use only nearby prices, but others (CRB) include other delivery months. A different index approach is taken by the Nasdaq/OMX Global Agriculture Index, which tracks the performance of shares in companies with activity in agriculture or farming. Since a commodity index cannot use market capitalization for a weighting scheme (all futures involve an equal long and short position), weights are determined by other means. The CRB uses equal weighting, but other indexes use world production or consumption data to reflect relative economic importance.

Changes in relative prices can alter the original weighting scheme over time, so an index may be rebalanced periodically. However, how an index is rebalanced can affect its performance since rebalancing can imply a particular investment strategy for the index portfolio. Rebalancing an equally weighted portfolio, for instance, necessitates selling appreciating commodities and buying depreciating commodities. Similarly, infrequent rebalancing can mimic a momentum investment strategy.

Diversification and Inflation

Because of the growing interest in alternative assets and their significant price appreciation, commodities are now considered an investable asset class. Further, since the factors determining commodity prices differ substantially from those affecting stocks and bonds, commodities are also increasingly attractive as diversifiers. Finally, commodities are used as inputs in manufacturing, and their prices play a significant role in manufacturing cost and output prices. Since the supply of many commodities is fixed because of long production lead times (in mining) or because of natural growing cycles (harvested crops), changes in demand can cause large change in prices, fueling inflation. For these reasons, commodities also may be a good inflation hedge.

Passive Investment Strategies

Positions in physical commodities involve large transactions costs, so many investors use more liquid alternatives, such as commodity futures, options on futures, or an index derivative. In the most common passive strategies, the investment involves a long-only futures position in individual commodities or a swap position based on an index combined with a simultaneous investment in cash. (Swaps are used because they allow for larger positions in the underlying futures.) Because only a small amount of funds is needed to obtain exposure to the commodity, the remainder of the investable funds is usually put in money market instruments.

The return from this strategy can be broken into various components. The *collateral return* is what is earned on the invested cash collateral, the real rate of return plus a return from anticipated inflation. The *spot return* is the price change on the commodity contract. This return would be positive, on average, if there is a risk premium. Alternatively, if spot price changes are driven solely by new information or by supply/storage factors, then the expected change in the spot price will be zero on average.

The *roll return* is derived from rolling the position forward as contracts mature, selling the nearby contract and reestablishing a long position in the same contract in a deferred maturity (rolling up or down the futures strip). In general, the roll return will be positive to the degree that futures prices exhibit backwardation. Backwardation is said to occur when the futures price is less than the current spot price, so that rolling the position forward involves selling the relatively higher-priced, maturing futures (which have converged to the spot price) and buying the lower-priced, deferred contract (whose price is less than current spot). Thus, the roll return depends on whether and how often backwardation occurs in the commodities contracts used. Roll return may be responsible for many empirical findings of a positive return (and presumably a positive risk premium) in commodity futures. Since contango (which occurs when the futures price is greater than the spot price) leads to negative returns, some index providers have adjusted their method of rolling futures contracts and take an active trading role in order to mitigate the effects of contango on index performance.

For an index strategy, a fourth return component reflects the *rebalancing* of index weights. If a portfolio of assets (in this case, commodity positions) is rebalanced, the portfolio will have higher returns than an unbalanced portfolio because of a reduction in variance. Further, this variance reduction is enhanced by assets with low correlation and high average variance, such as commodities. This means that a commodity index strategy will have a rebalancing return that will not be found in an investment in individual commodities. The size of this rebalancing return is sensitive to when and how frequently the index is rebalanced.

Recently, additional versions of commodity indexes have been introduced that separate out components of the return on a commodity futures portfolio. For instance, the DJ-AIGCI family of indexes includes a Total Return subindex that reflects the return on collateralized positions in futures, and differs from the excess return-DJ-AIGCI that tracks futures price movements only. In addition, a third index, the DJ-AIG Spot index, measures changes in commodity *spot* prices and does not reflect rollover or collateral effects.

A simple example of a long speculative position in gold, and the effect of the roll return component, is given in Exhibit 6.1. At the beginning of August, a speculator forecasts an increase in the price of gold over the next month due to increasing risk aversion in the United States. To take advantage of rising gold prices, the speculator buys 10 futures contract on gold.

Exhibit 6.1 Long Speculation in the Gold

Futures Prices		
Date	**Futures Maturity**	**Settlement Price**
1-Aug	Aug 07	663.60
	Sep 07	664.90
15-Aug	Aug 07	668.90
	Sep 07	670.70
31-Aug	Sep 07	673.00

Alternative A:

Buy 10 September futures on August 1 and hold until August 31:
Futures gain = (\$673.00 – \$664.90) × 100 oz/contract × 10 contracts = \$8100.00

Alternative B:

Buy 10 nearby August futures and roll over into September futures on August 15:

August futures: (\$668.90 – \$663.60) × 100 × 10 =	\$5300.00
Sept. futures: (\$673.00 – \$670.70) × 100 × 10 =	\$2300.00
Futures gain	\$7600.00

The speculator can choose to buy the September futures contract at \$664.90 per oz. and hold that position until the end of August, when she sells the contact back at \$673.00. With this approach, she will make a profit of \$8100.00 (= 10 contracts × 100 ounces/contract × (\$673.00 – \$664.90)). Alternatively, she could buy the nearby August contract and then roll her position into the next delivery month (September) in mid-August when the August contract expires. With this approach, she has two sets of transactions. She buys the August contract on August 1 at \$663.60 and sells it back just before its maturity on August 15 at \$668.90. To maintain the long position, she simultaneously buys the September contract at \$670.70. At the end of the month, she sells the September contract back at \$673.00.

In either case, the speculator will gain from her long position as gold prices rise through the month. However, by closing out the August futures and reentering the market at the higher September futures price, rolling up the futures strip results in a negative roll return and illustrates the effect that the roll return can have on a long position when prices are in contango rather than backwardation.

Other passive investment alternatives are available. Exchange-traded funds (ETFs) give the buyer an equity participation in the performance of individual commodities, sectors, or index. Deutsche Bank and PowerShares offer a number of commodity-based ETFs tracking precious metals such as gold and silver, base metals, and agriculture. Similarly, Barclays Bank offers ETFs (iShares) based on the SP-GSCI, gold, and other commodities.

An exchange-traded note (ETN) involves a third-party–issued debt instrument in which the returns on the bond are linked by an agreed formula to the performance of a basket of commodities or a commodity index. Such notes expose the holder to an additional risk, the credit risk of the issuer. Last, an investor can also acquire the stocks of natural resource firms as a proxy for an investment in commodities, but this approach includes exposure to additional risks related to firms' particular management, strategy, and risk management policies.

Active Strategies

An actively managed exposure to commodities can be acquired through an investment in a managed futures fund or a commodity pool managed by a commodity trading advisor (CTA), or in a commodity-based mutual fund. Managed futures are similar to hedge funds although restricted to investment in futures only. Some managed futures funds limit their exposure to a specialized sector, such as metals, or agricultural commodities. Commodity mutual funds specialize in commodity-related investments and are very similar to an investment in a financial asset mutual fund. The best-known commodity mutual funds are offered by Oppenheimer and PIMCO.

In all active strategies, value is created by the fund manager. Generally, two types of strategies are followed in such funds. Trend followers use fundamental and technical analysis signals to go long or short in futures markets. Market-neutral managers use any number of value-driven strategies, including spreads between related markets and options strategies to trade on relative movements in commodity prices. As with financial funds, active commodity funds are compared to one of the commodity indexes to measure the performance of the manager's particular investment choices.

Measuring Investment Performance

Empirical tests of the performance of commodities have examined several issues, including comparisons of risk and return versus other asset classes and their use as an inflation hedge and portfolio diversifier. The results concerning the historical price behavior for commodities compared to stocks, bonds, Treasury bills, and other asset classes have been mixed. Empirical results are affected by the choice of commodities and time period, whether individual or portfolio returns are used, and the weightings used to construct commodity portfolios. While some studies have found that a commodity portfolio exhibits returns commensurate to a well-diversified stock portfolio (Bodie and Rosansky 1980; Gorton and Rouwenhorst 2006), returns can vary widely across commodities, the business cycle, and monetary policy regime (Kat and Oomen 2007a). Further, the weighting scheme and rebalancing choices can affect the rate of portfolio return significantly over time.

The results concerning the performance of actively managed funds are also mixed. Managed futures performance seems to depend on the time period examined: While Irwin and Landa (1987) find that commodity pools exhibited good returns in the 1970s, managed futures were poor performers in the 1980s (Edwards and Park 1996).

There is some evidence that commodity returns vary directly with the degree of backwardation. The size of price backwardation in a particular commodity futures strip determines the size of the roll return, which, absent a positive risk premium, is the only return source for the investment besides the collateral return (Erb and Harvey 2006). Backwardation is itself related to the size of the convenience yield embedded in the futures price (Feldman and Till 2007). While convenience yield is associated with commodities that are hard to store, its size, or even whether backwardation occurs, varies considerably through time. Thus, choosing the components of the commodity portfolio may involve forecasting convenience yield and the size of price backwardation. Indeed, as Erb and Harvey (2006) point out, the two indexes most commonly used in commodity strategies (SP-GSCI and DJ) are heavily weighted in the energy sector, and their recent positive performance may be explained by the fact that these commodity futures have exhibited backwardation more frequently than other commodities.

Are commodity futures a good inflation hedge? In general, commodities are positively correlated to inflation measures (Bodie 1983), and futures perform better than stocks in years of high inflation. Both Greer (2002) and Gorton and Rouwenhorst (2006) find that a commodity portfolio was positively correlated to inflation and changes in inflation. Fama and French (1987) found that the basis in many commodities, particularly metals, was positively related to nominal interest rates. Also, commodities futures perform well in the early stages of a recession but have lower returns in the later stages of a recession, showing that some of the negative correlation of commodities with equity returns relates not only to inflation but also to the differing behavior of equities and commodities during the business cycle. However, Erb and Harvey (2006) found that the relationship of returns to inflation and inflation changes varies considerably across individual commodities, and those commodities that exhibit roll returns (usually associated with storage difficulty) perform well during periods of unexpected inflation.

In general, commodities appear to be an attractive diversifier. Most studies have found a low or negative correlation between commodities and security returns (Jensen, Johnson, and Mercer 2000), but the correlations are unstable and vary significantly between commodity sectors (Kat and Oomen 2007b). In addition, the diversification benefit of commodities can be counterbalanced by the reduction in portfolio return resulting from their addition to a portfolio.

HEDGING

Hedging is usually associated with risk management, the desire of producers or consumers to reduce the risk of price changes. As producers, farmers want to protect against crop price declines and will go short futures (or buy put options), as will metals producers. Processors and exporters, however, want to hedge against an increase in purchase prices, and go long futures (or buy a call option). Traditional commercial commodity hedging involves profit maximization in addition to pure price risk management. These marketing programs choose the optimal time and price at which to market their crop using their local basis.

Commodity Marketing

A common operational decision for a commodity purchaser (one who needs the commodity as a future input in a manufacturing process) is whether to enter a cash forward purchase at a particular price or wait and use derivatives to hedge until a forward purchase is made at a better price. A similar decision process, but with reverse positions, is involved with a commodity seller.

By going long in the futures market, the purchaser expects to fix a purchase price that will equal the futures price plus the expected basis at maturity. This approach to the hedging decision differs from the traditional assumption of a perfect hedge, with no change in the basis and a final purchase price equal to the futures price at hedge initiation. Note that a weakening basis, in which the futures price rises more than the spot price (or falls less than the spot price) will result in a gain on a long hedge: The gain on the long futures position is greater than the loss on the spot price to be paid, or, alternatively, the loss on the futures price is less than the gain on a lower purchase price. In this framework, the hedger uses knowledge of basis patterns to predict a weakening basis and choose when to fix the best price.

As an example of a traditional approach to marketing management, assume it is June, and a miller needs to purchase wheat for November delivery. The futures contract maturing after the anticipated delivery date is usually used to avoid lifting the hedge before cash delivery, so the reference futures price he uses would be the December futures contract. Considering the cash forward offers for November delivery from grain elevators of $4.05, $4.10, and $4.15 per bushel, the miller calculates his basis (cash forward offer less December futures price of $3.90) for each of the offers, which are 15 cents, 20 cents, and 25 cents over, respectively. Note that a negative basis—for instance, a corn cash price of $4.50 less the futures price of $4.75—is referred to as 25 cents under (cash price is "under" the futures price). A positive basis is referred to as over. Based on historical data for the December futures in previous Novembers, the expected, or average, basis as of November is 10 cents over. Previous experience shows then that the basis on the December contract can be expected to weaken by November. Because a weakening basis will result in a positive cash flow for the long hedger, the commodity purchaser uses this as a signal to wait to make a purchase until the basis weakens as he expects. To hedge against price volatility in the meantime, however, he buys December futures contracts at $3.90, entering into a long hedge.

If the basis does weaken as expected, the miller will then be able to enter a forward cash purchase at a much better overall price. To illustrate, in October, the processor returns to the cash forward market, getting offers of $4.15 and $4.20 (basis of 5 cents and 10 cents over, respectively, to the December futures price of $4.10). The processor decides to purchase at the cash forward offer of $4.15, which has a weaker basis (5 cents over) than the average, expected basis of 10 cents over. Because he postponed his cash forward purchase and hedged in the futures market, the processor has fixed a more attractive price than was available to him in June. Although the cash forward price is higher now, at $4.15 compared to June's offer of $4.05, the weakening of the December futures basis to 5 cents over means that the hedge has made him more than enough to lower the overall purchase price. The processor went long December futures at $3.90 and sold at $4.10, making 20 cents on the hedge. The total purchase price, then, is $3.95: $4.15 cash price less $0.20 futures return = $3.95. In this example, the purchaser used his knowledge of basis behavior to time his purchase in the cash forward market and to choose when to put on an anticipatory hedge in the December futures.

This example is a traditional approach to marketing management. The availability of options on commodity futures has increased the flexibility of risk management programs to deal with changing conditions. For instance, for a processor (long hedger), a basic hedge strategy involves a pure long futures or long call hedge. But in a time of falling prices, the long hedger can use additional call options, or put options, to restructure the hedge to manage this price movement. If prices fall unexpectedly, a hedger who purchased a call option can move her ceiling price by purchasing an additional call option with a lower strike price. With this second call, the producer can avoid lifting the hedge in response to falling prices while capturing some of the potential purchase price decline through the lower strike price (at the cost of the premium paid for the additional option). Alternatively, if prices fall, a long hedger could then purchase a put option. This creates a synthetic call and allows the farmer to capture some of the price change without, again, having to lift the hedge.

Risk Management

As with financial futures, two key issues in commodity risk management involve deriving an optimal hedge ratio and estimating the ratio accurately. The hedger usually is assumed to minimize risk, generally specified as return variability, but alternative risk measures include mean-Gini coefficient or generalized semivariance. Some models instead maximize utility of hedged returns, although under some assumptions on the futures price, the hedge ratio under the two different objectives will be the same.

An example of a long hedge in corn is given in Exhibit 6.2. In September, a food processor forecasts that he will need to purchase 500,000 bushels of corn at the end of October. Given the strong upward trend in corn prices, the processor is concerned

Exhibit 6.2 Long Hedge in Corn

Date: September 14
Current spot price = \$3.305/bushel
December futures price = \$3.49/bushel

Buy 100 Dec futures contracts at FP = \$3.49 per bushel

Value of futures position:	Value of spot position:
100 contracts × \$3.49/bu × 5,000 bu/contract = <u>\$1,7450,000</u>	500,000 bu × \$3.305 = <u>\$1,652,500</u>

Date: October 31
Current spot price = \$3.59/bushel
December futures price = \$3.755/bushel

Sell 100 Dec futures contracts (offset hedge) at FP = \$3.755 per bushel

Value of futures position:	Value of spot position:
100 contracts × \$3.7550/bu × 5,000 bu/contract = <u>\$1,877,500,</u>	500,000 bu × \$3.59 = <u>\$1,795,000</u>

<u>Results</u>:
Gain = \$132,500 | Loss = \$142,500
Net loss = \$10,000

that spot prices for corn will be much higher than the current spot price of $3.305 per bushel. The processor decides to hedge the anticipated purchase by buying (going long) 100 December corn futures contracts at a futures price of $3.49 per bushel. Since each corn futures contract calls for delivery of 5,000 bushels, the hedge ratio (the amount of the commodity underlying the futures position as a percentage of the cash commodity hedged) is 1.0, with 500,000 bushels in futures (= 100 contracts × 5,000 bushels/contract) to hedge the 500,000 bushel anticipated cash purchase.

At the end of October, the processor is ready to make his corn purchase and closes out his hedge by offsetting his futures position, selling 100 December futures at the futures price of $3.755, making a profit of $132,500 on his position. While the processor was correct in his forecast of a cash price increase and must now purchase cash corn at $3.59 per bushel, a $142,500 increase over the September price, the hedge proceeds cover almost all of this increased cost.

Looked at in another way, the change in the futures price of 26.5 cents per bushel (= $3.755 – $3.49) offsets almost all of the increase in cash corn prices of 28.5 cents per bushel (= $3.59 – $3.305). With the hedge, the processor has fixed a total price of $3.325 for his cash corn purchase ($3.59 cash price in October less the hedge proceeds of $0.265 = $3.325/bushel), which is very close to the cash price for corn at the start of the hedge in September of $3.305 per bushel.

An example of a short hedge in gold, with an optimal hedge less than 1.0, is given in Exhibit 6.3. At the end of March, a small gold mining company is concerned that gold prices will fall from their current level of $675.55 per ounce before it can sell its production of 10,000 ounces of gold in two months. It hedges by selling 80 July gold futures contracts at a futures price of $683.70 per ounce. In this example, the hedger decides to use a more sophisticated approach to hedging than

Exhibit 6.3 Short Hedge in Gold

Date: March 30 Current futures price July futures $683.70	Risk-minimizing OLS hedge ratio: Regression Results: Δ spot price $= \alpha + \beta\ \Delta$ futures price $+ \varepsilon$ Hedge ratio $= \beta = 0.8043$
Futures position: March 30: sell 80 July futures contracts at FP = $683.70 per oz.	**Spot position:**
Value of futures position:	Value of spot position =
80 contracts × $683.70/oz × 100 oz/contract = $5,469,600	10,000 oz × $675.55 = $6,755,500
May 31: buy 80 July futures contracts at FP = $650.90 per oz.	
Value of futures position:	Value of spot position =
80 contracts × $650.90/oz × 100 oz/contract = $5,207,200	10,000 oz × $649.50 = 6,495,000
Gain = $262,400 Net gain = $1,900	Loss = $260,500

a simple one-to-one hedge ratio. Using past gold spot and futures prices, the basic risk-minimizing ordinary least squares (OLS) model results in an optimal hedge ratio of 0.8043: The optimal hedge requires a futures position equal to 80.43 percent of the cash position of 10,000 ounces, or 8,043 ounces. The miner sells 80 contracts, or 8,000 ounces of gold in the futures market (= 80 contracts × 100 oz./contract). With gold, the hedger has a choice of contract delivery months of May, June, or July. July is chosen rather than May so that it will not be necessary to roll over a futures position due to contract expiration and also because the July contract will have more trading volume at the end of the hedge than the soon-to-expire June contract.

By the end of May, gold prices have decreased considerably, falling to $649.50 per ounce. The miner lifts the hedge, buying 80 July gold futures contracts at a price of $650.90 per ounce. The proceeds from the hedge of $262,400 (= 80 contracts × ($683.70 – $650.90/oz) × 100 oz/contract) more than make up for the miner's loss in the cash market of $260,500 (10,000 ounces × ($675.55 – $649.50)). Note also that the risk-minimizing optimal hedge model provided a better estimate of the appropriate hedge ratio (80.43%) than a simple one-to-one approach.

Because it is easy to estimate empirically, the risk-minimizing OLS model is often used to calculate a hedge ratio. A wide array of statistical adjustments to this basic model has been suggested to improve its performance. For instance, a form of MGARCH modeling is used to allow for time-varying volatility from seasonals or time to maturity (Baillie and Myers 1991; Sephton 1993). Other statistical adjustment methods include cointegration, cointegration-heteroscedastic, or random coefficient methods (Lien and Shrestha 2007).

Since Metallgesellschaft experienced heavy losses due to maturity mismatching in 1993, there has been renewed interest in dynamic and multiperiod hedge models beyond the basic choice of a stack or strip hedge. Multiperiod models tend to have substantial data requirements; an alternative approach is to adjust the minimum variance hedge ratio periodically (Lence, Kimle, and Hayenga 1993). Last, some multiperiod hedging models base their adjustment process on models of commodity price dynamics, particularly the switching behavior between investment asset (full cost of carry) and convenience commodity pricing.

Commodity hedgers can also be subject to multiple sources of risk beyond price volatility alone (Haigh and Holt 2000). Crop or production *yield* can be uncertain, causing both price and quantity risk. Further, since commodities require physical delivery, variability in transport costs can also be a source of risk. With multiple risks, the hedging decision is complicated because of interrelations among price, amount produced, and transport cost at harvest or sale. In addition, commodity producers are also subject to "own basis" risk because the price used to settle the hedge instrument usually is not the same as the price for the producer, given regional and quality differences. Thus, commodity hedges frequently involve several sources of uncertainty in related but distinct markets.

SPREADS

Some commodities' prices are linked through a common production process. Prices for soybeans and its products, soy oil and soy meal, are linked by anticipated production costs, as are feeder cattle, corn, and live cattle. Such relationships give rise to spread trade opportunities.

Spreads involve a simultaneous long and short position in related futures contracts in anticipation of a relative price movement. Because the spread involves two related positions, it involves lower risk, lower margin, and lower return. Generally, the price relationship is relatively stable and in many cases mean-reverting. Interdelivery, or calendar spreads, involves a long and short position in different delivery months for the same commodity. Spreads of this kind are based on expected changes in carry charges between months. Interexchange spreads, with long and short positions in the same contract traded on different exchanges, derive from transportation costs expectations. Intercommodity spreads depend on the expected price correlation between related commodities. Soybean prices are related to soy oil and soy meal through the cost of refining the two products from beans. Similarly, feeder cattle or live hogs are related to the prices of feed (corn but also soybeans) and final product, live cattle or pork bellies, respectively.

Both hedgers and speculators use spreads. Hedgers will use a spread to lock in a profit margin or to manage operating risk. Speculators usually enter a spread when they perceive a misalignment between current price spreads and historical or expected processing costs.

CONCLUSION

Speculating and investing in commodity futures can be either passive, usually through a commodity index, or active, and commodities also can use as a diversification tool or as an inflation hedge in portfolio investment strategies. Risk management can include traditional marketing programs, in which futures price behavior is used to time transactions and maximize profit, and hedging, in which futures are used to transfer risk. Spreads can be used by both hedgers and speculators. See also Chapters 3 and 9.

REFERENCES

Baillie, R. T., and R. J. Myers. 1991. "Bivariate GARCH Estimation of the Optimal Commodity Futures Hedge," *Journal of Applied Econometrics* 6: 109–124.

Bodie, Z. 1983. "Commodity Futures as a Hedge against Inflation," *Journal of Portfolio Management* (Spring): 12–17.

Bodie, Z., and V. I. Rosansky. 1980. "Risk and Return in Commodity Futures," *Financial Analysts Journal* (May-June): 27–39.

Edwards, F., and J. Park. 1996. "Do Managed Futures Make Good Investments?" *Journal of Futures Markets* 16: 475–517.

Erb, C. B., and C. R. Harvey. 2006. "The Strategic and Tactical Value of Commodity Futures," *Financial Analysts Journal* (March-April): 69–97.

Fama, E. F., and K. R. French. 1987. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *Journal of Business* 60: 55–73.

Feldman, B., and H. Till. 2007. "Separating the Wheat from the Chaff: Backwardation as the long-term Driver of Commodity Futures Performance. Evidence from Soy, Corn and Wheat Futures from 1950 to 2004," Premia Capital and Prismatics Analytics White Paper (SSRN id954874).

Gorton, G., and K. G. Rouwenhorst. 2006. "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal* (March-April): 47–68.

Greer, R. J. 2000. "The Nature of Commodity Index Returns," *Journal of Alternative Investments* (Summer): 45–53.
Haigh, M. S., and M. T. Holt. 2000. "Hedging Multiple Price Uncertainty in International Grain Trade," *American Journal of Agricultural Economics* 82: 881–896.
Irwin, S., and D. Landa. 1987. "Real Estate, Futures, and Gold as Portfolio Assets," *Journal of Portfolio Management* (Fall): 29–34.
Jensen, G. R., R. R. Johnson, and J. M. Mercer. 2000. "Efficient Use of Commodity Futures in Diversified Portfolios," *Journal of Futures Markets* 20: 489–506.
Kat, H. M., and R. C. Oomen. 2007a. "What Every Investor Needs to Know about Commodities I: Univariate Return Analysis," *Journal of Investment Management* 5: 1–25.
Kat, H. M., and R. C. Oomen. 2007b. "What Every Investor Needs to Know about Commodities II: Multivariate Return Analysis," *Journal of Investment Management* 5: 16–40.
Lence, S. H., K. L.Kimle, and M. L. Hayenga. 1993. "A Dynamic Minimum Variance Hedge," *American Journal of Agricultural Economics* 75: 1063–1071.
Lien, D., and K. Shrestha. 2007. "An Empirical Analysis of the Relationship between Hedge Ratio and Hedging Horizon Using Wavelet Analysis," *Journal of Futures Markets* 27: 127–150.
Sephton, P. S. 1993. "Optimal Hedge Ratios at the Winnipeg Commodity Exchange," *Canadian Journal of Economics* 26: 175–193.

SUGGESTED FURTHER READING

Carter, C. A. 1999. "Commodity Futures Markets: A Survey," *Australian Journal of Agricultural and Resource Economics* 43: 209–247.
Lien, D., and Y. K. Tse. 2002. "Some Recent Developments in Futures Hedging," *Journal of Economic Surveys* 16: 357–396.

ABOUT THE AUTHOR

Joan C. Junkus is an associate professor of finance at DePaul University in Chicago. She received both her MBA and PhD from the University of Illinois in Urbana. The author of numerous articles on futures pricing and hedging in such journals as the *Journal of Futures Markets* and *Derivatives Use, Trading & Regulation*, she serves on the board of *Managerial Finance*. Dr. Junkus is also the coauthor of the textbook *The Theory and Practice of Futures Markets*. Currently pursuing her primary interest in the history of financial markets and the development of alternative investments, she has recently published in *Financial History*. Dr. Junkus has served as associate dean of the College of Commerce at DePaul as well as chair of the Finance Department. Earlier, Dr. Junkus was on faculty in the Graduate School of Management at the University of California-Irvine and worked for Salomon Brothers in Chicago.

CHAPTER 7

Equity Derivatives

JEFFREY H. HARRIS
Professor of Finance, University of Delaware, and Chief Economist, United States Commodity Futures Trading Commission

L. MICK SWARTZ
University of Southern California

INTRODUCTION

News about the stock market dominates the popular press. Since investors make stock investments based on their expectations about the future, stock market returns reflect investor expectations. For individual companies, however, the future may hold great uncertainty about firm performance. This uncertainty represents the risk of investing in a company. Derivative securities have been designed to manage these risks. Stock options, single stock futures, convertible bonds, warrants, and equity swap agreements can all be used to manage firm-specific risks.

While developing portfolio theory, Harry Markowitz recognized that building portfolios of stocks can help to manage risk as well. The correlations between stocks in a portfolio can serve to reduce overall portfolio risk, even while individual companies remain risky. In this regard, portfolios can help to reduce firm-specific risk. However, without perfectly negative stock correlations, every portfolio also contains systematic risk that is not diversifiable. This portfolio risk also can be managed with a set of derivative securities. Options, futures, and swap products each help to manage portfolio risk. Among these, tailored swap contracts have experienced the most dramatic growth in recent years while futures and options products remain the most popular equity portfolio derivatives.

According to Futures Industry Association (FIA) figures, global options and futures trading volume rose by 28 percent in 2007 (Burghardt 2008). This growth resulted in more than 15 billion futures and options contracts traded that year. By any standard, global growth in these markets has been robust for the past five years with annual growth rates of 30 percent in 2003, 9 percent in 2004, 12 percent in 2005, and 19 percent in 2006. FIA figures show that equity derivatives accounted for 64 percent of the increase in futures and options trading volume.

We have benefited greatly from the comments of Jim Overdahl. The views expressed in this chapter are those of the authors and should not be viewed as those of the U.S. Commodity Futures Trading Commission or any of the commissioners.

Increases in worldwide trading volume have largely mirrored the increase in volatility of the underlying equities markets. Market participants around the world seeking to hedge against volatility in the underlying equity markets are increasingly drawn to the equity derivatives markets. For instance, Asian demand for derivatives grew substantially in 2007. Trading volume on the Korean exchange increased by 9 percent during 2007. More strikingly, the Hong Kong exchange reported an increase in derivatives trading volume of more than 100 percent during 2007. Elsewhere in the world, growth rates have been even more dramatic with 214 percent growth in South Africa, for instance.

From a worldwide perspective, increased trading volume mirrors the growth in contract originations. Although approximately 40 percent of derivatives originated in North America, derivative contracts are truly international in scope, with another 28 percent of contracts originated in Asia, 22 percent in Europe, and 7 percent in Latin America. As emerging nations develop financial systems, these countries are increasingly using derivatives to manage risk. Although volatility of the underlying assets drives much of the growth worldwide, some of the increase is due to more market participants using derivatives.

Given the growing array of derivative choices—options, futures, swaps, and their various combinations—the choice of appropriate instruments can be difficult. In a 2008 Tabb survey (Nybo 2008) of financial institutions, liquidity considerations were rated as the top concern of asset managers when picking from the set of available derivative instruments. A full 70 percent of institutions listed liquidity as the biggest challenge in derivatives markets. In contrast to many other derivatives markets, nearly all equity derivatives markets have been developing greater liquidity over time. In response to greater comfort with liquidity levels, a majority (53 percent) of asset managers have relaxed their internal restrictions on trading options.

STOCK OPTIONS

Stock options represent risk management tools for individual company investments. As risk management tools, stock options typically are traded most actively when firm risk is high. It is unusual, for instance, to see stock options on utility companies, whose regulated status makes for largely predictable returns. Stock option trading flourishes in companies engaged in technology development, biotechnology and other higher-risk ventures. Where uncertainty reigns, stock options can serve to mitigate the risk of stock investing. Indeed, as the idiosyncratic risk of individual stocks has been increasing over time, growth in stock option trading has expanded concurrently.

Options on individual stocks are, in fact, some of the most popular of the equity derivatives. Stock option trading volume, although driven by volatility, is tempered by liquidity considerations. Indeed, liquidity was identified as the greatest challenge for institutional investors during 2008, particularly for options on small-capitalization stocks.

Technology, however, is playing an increasingly central role in providing liquidity sources to options trading. Liquidity in equity options markets is accessed across a variety of venues, including directly from traders on an organized exchange, through over-the-counter (OTC) option contracts, through direct access via electronic market platforms, and through smart order routers that can access many of these venues automatically. As a result, sell-side institutions continue to design more complex and functional electronic systems that allow for greater access to and aggregation of liquidity.

The complexity of stock option pricing slowed the development of electronic trading in these markets. While Nasdaq pioneered electronic equity trading starting in the early 1970s, it was not until 2000 when the electronic International Securities Exchange (ISE) brought serious electronic competition to options trading. Since 2000, the ISE has captured a significant market share in U.S. stock option trading. Just as Nasdaq has grown to trade more than 2 billion shares per day electronically, by 2010 approximately two-thirds of all stock option trading will be done electronically as well.

The mechanics of options contracts are quite simple. Option contracts involve both a buyer and a writer. The buyer buys a form of insurance and the writer provides the insurance and is obligated by the terms of the contract. The buyer pays a premium to the writer at the beginning of the contract. The writer is then obligated to buy or sell the stock to the option buyer at a specified price during the length of the contract. Options traded on exchanges have rules requiring the writers of these contracts to show they can honor their contracts on a daily basis. Since the buyer pays the entire premium up front for these option contracts, there is no need to ensure the buyer can pay.

An American option allows the buyer of an option to buy or sell the stock on or before the exercise date. A European option allows the buyer of an option to buy or sell the stock only on the exercise date. As a result, an American option with all other terms equal is worth at least as much as a European option.

Call Options

Stock option prices are affected by six variables: the stock price, the exercise price of the option, the time to expiration, the risk-free rate, the volatility and expected dividends during the life of the option. If the price of a stock rises, the value of a call option rises. Specifically, the value of a call option on the expiration date is

$$V(\text{call}) = \max(0, S - X)$$

where *V(call)* = *value of the call*
S = stock price on the expiration date
X = exercise price on the expiration date

If the price of a stock rises, the value of a call option rises. The call option provides protection from an increase in the value of the asset. If the exercise price is lower, the value of the call option is higher.

In addition to the stock price and the exercise price, the value of a call option is affected by the time to expiration, the volatility of the stock, the risk-free rate of interest, and the expected dividend of the stock. The value of a call option rises if the risk-free interest rate increases and all other factors remain unchanged. The impact of time affects American options differently from European options. An increase in time to expiration increases the value of all American options (puts and calls). Investors can choose to exercise their option among more time periods, encompassing more states of the economy. As a result, the value of an American option rises with an increase in the time to expiration.

Since a European option can be exercised only on the expiration date, increasing the time to expiration may not increase European call values. Another factor affecting the value of a stock option is the dividend. Since dividends reduce the value of a stock, larger dividends reduce call values.

Put Options

The value of a put option on the expiration date is

$$V(\text{put}) = \max(0,\ X - S)$$

where *V(put) = value of the put*
$S =$ stock price on the expiration date
$X =$ exercise price

The value of a put option increases as the value of a stock decreases. A put option provides protection from a decrease in the price of an asset. Like call, put options increase in value with an increase in volatility and an increase in the time to expiration (for American options). A put option decreases in value as the risk-free interest rate in the economy rises (opposite of the call option). The impact of the value of dividends distributed during the life of a put option is opposite the impact of a call option since dividends reduce stock prices.

Two commonly used models for pricing options prior to expiration are the binomial model and the Black-Scholes model. These models can be used to demonstrate that options prices are very sensitive to volatility levels and where volatility is a concern options can be used for effective hedging.

Stock Options on an Index

A Tabb Group survey of financial institutions published in 2008 estimates that approximately 30 percent of institutional investors actively use options to gain exposure to or hedge the risk, with stock index options representing the most popular of all options. This percentage is expected to double in the next few years. In fact, financial institutions have made equity index options some of the most actively traded contracts in the world. On the basis of trading volume, the largest 10 equity index futures and options contracts worldwide are:

(1) Kospi 200 options
(2) E-mini S&P 500 Futures
(3) DJ Euro Stoxx 50 Futures
(4) DJ Euro Stoxx 50 Options
(5) Powershares QQQQ ETF options
(6) Standard & Poor's (S&P) 500 options
(7) iShares Russell 2000 ETF Options
(8) SPDR S&P 500 ETF Options
(9) S&P CNX Nifty Futures (India)
(10) E-mini Nasdaq 100 Futures

In the United States, the S&P 500, Dow Jones Industrial Average, Nasdaq 100, and S&P 100 are the most heavily traded index options.

Stock index options are settled in cash. This feature simplifies the process at settlement, since no delivery of the underlying index is required. The relative simplicity at settlement also contributes to the popularity of index options as hedging instruments. For optimal or partial portfolio hedging with index options, the number of put options to hedge is proportional to the portfolio beta.

Employee Stock Options

Many employees receive stock options, which provide incentives to employees to take actions that benefit the firm. Top managers make key decisions that influence the value of the firm. By receiving a form of compensation that increases with the stock value, employees have incentive to take actions that are of benefit to other shareholders. In contrast to typical call options, employee or executive stock options are long-term call options on the stock—many last 5 to 10 years and commonly they are not exercisable for at least 3 to 5 years. Employee options cannot be traded on exchanges (are illiquid), and new shares are issued at the time of exercise, diluting the value of existing shareholders. Furthermore, employees can take actions that affect the value of the option. For these reasons, employee stock options are more difficult to price than standard stock options.

Convertible Bonds

Although not originating as equity, debt issued as convertible bonds holds the potential to convert into stock under specific terms in the bond contract. In this regard, convertible bonds contain a derivative component related to the equity value. The holder of a convertible bond effectively holds a contract with a fixed bond component and a call option on shares of the firm. When the call component is out of the money, the convertible bond simply remains debt on the company balance sheet. However, if the call component is in the money, the bond can be converted into shares of stock.

As with a call contract, holders have no incentive to exercise the conversion option early—early exercise destroys the time value of the option. The optimal strategy for holders would be to sell the convertible bond to another investor and buy shares with the proceeds to reap the inherent time value of the option. However, the issuer typically would like holders to convert early, so many convertible bond contracts contain call provisions. Call provisions in convertible callable bonds give both the bond issuer and the bondholder options: The issuer has a call option on the underlying bond and the holder has a call option on the firm's stock. Although the holder receives no benefit from early exercise, the bond issuer does benefit from forcing conversion. By forcing conversion, the bond issuer eliminates the time value embedded in the conversion option.

One difference between the value of the convertibility option and a simple equity option is the fact that conversion of the debt to stock leads to an increase in the company shares outstanding (e.g., employee stock options), diluting share value and creating a more complex pricing problem for the convertible debt at issuance.

Warrants

A warrant is a long-term call option that allows a warrant holder to buy shares directly from the firm at the exercise price. A small company might issue stock at $15 per share and allow stockholders to buy warrants for $1 per share. The warrant may allow buyers to purchase the stock at $30 per share anytime over the next five years. If the stock price rises to $40 per share within the next five years, the warrant holder could buy the stock at $30 per share (the company issues more shares so dilution occurs) and could sell the shares for a $10 profit on a $1 investment. As with convertible debt, the option to exercise a warrant also has time value, so early exercise is not likely. Since early exercise destroys the time value of the option, a warrant holder would always be better off selling the warrant and buying the underlying stock rather than directly exercising the warrant for shares.

As with convertible bonds, the exercise of warrants forces the firm to issue new shares to the warrant holder, increasing the shares outstanding for a company. Exercised warrants also dilute share values. If the value of the firm never rises above $30 per share over the five-year time period, the firm just keeps the $1 premium for the warrant. Therefore, the firm raises cash at the expense of possibly sharing future gains with the warrant holders.

EQUITY FUTURES

Equity futures exist for both single stocks and portfolios of stocks. Worldwide, forward trading in stocks dates back more than a century in Swiss markets. Active equity futures markets in Europe, Asia, and Australia have existed for more than 20 years, with equity index futures representing the majority of trading volume. Although relatively lightly traded, single-stock futures have been traded in Sweden and Finland for more than a decade.

While stock index futures have been traded in the United States for more than three decades, single-stock futures and narrow-based index futures became legal to trade in this country only after passage of the Commodity Futures Modernization Act (CFMA) in 2000. The CFMA ended an 18-year government ban on single-stock futures and narrow-based index futures, a ban that stemmed largely from regulatory jurisdictional disagreements. The combination of security (equity) features regulated by the U.S. Securities and Exchange Commission (SEC) and futures characteristics regulated by the Commodity Futures Trading Commission (CFTC) created a quandary that led to the effective ban. Although single-stock futures are designated as futures contracts, concerns about potential manipulation of the underlying stock(s) raised concerns from the SEC.

Single-Stock Futures

Single-stock futures have been developed as an alternative for managing the risk of investing in stocks. Single-stock futures offer a market to buy or sell the underlying stock at some future date, typically at some short horizon (within a year). They represent a somewhat lower-cost alternative to stock options, since entering into a futures contract requires only margins. Worldwide, and most predominantly in

South Africa, single-stock futures trade on a wide variety of stocks. As with stock options, however, trading is likely to be concentrated in firms where uncertainty is greatest.

Although in the United States, single-stock futures typically are settled on a $T+3$ basis with the delivery of the underlying stock, the use of single-stock futures facilitates short positions in stock. By using single-stock futures, an investor avoids the costly logistics required for borrowing shares in a short sale arrangement. Further, the relatively low (20 percent) margins on U.S. single-stock futures make for a low-cost competitive alternative to options for hedging purposes.

Single-stock futures are priced with a simple present value/future value relation. More specifically, the price of a single-stock future should represent the present value of today's stock price along with the present value of any future dividends that might be paid out before the futures contract expires. Mathematically, we can state that the value of a single-stock future F is

$$F = (S_o - D)e^{rT}$$

where S_o = current stock price
D = present value of dividends paid during the life of the futures contract
T = time to maturity (in years)
r = continuously compounded risk-free rate of interests over the life of the futures contract

Futures on Stock Indexes

Some of the most actively traded futures contracts involve stock index futures. Stock index futures range from broad-based to narrow-based indexes. The most broadly traded index futures are based on country-specific indexes such as the S&P 500, the Financial Times Stock Index FTSE 100 (London-listed companies), the Deutschen Actien Index DAX 30 (Frankfurt-listed companies), and the Cotation Assistee en Continu CAC 40 (Euronext Paris-listed companies). Stock indexes based on other countries typically are narrower in scope. For instance, the Portugese Stock Index PSI-20 index includes 20 Euronext Lisbon-listed companies, with the top 5 firms representing approximately 75 percent of the market capitalization of the entire index.

In the United States, OneChicago lists a number of narrow-based index futures. The indexes traded here are typically comprised of portfolios of four to seven stocks, many slanted toward Canadian stocks.

Portfolio managers at mutual funds, hedge funds, insurance companies, and other institutions face systematic portfolio risks when they hold equity portfolios. Likewise, market makers and dealers who sell index products to clients also can be exposed to systematic risk. Stock index futures provide an efficient mechanism for managing this portfolio risk at a relatively low cost. Since entering into a futures contract requires only margin payments, these institutions can avoid the premium payment that accompanies index options.

EQUITY SWAPS

The equity swap market has emerged from the need for more tailored derivatives. Equity swaps, as the term suggests, involve two parties that swap payments based on the notional value specified as some portfolio of equities. The most basic equity swaps involve fixed-for-floating payments based on the underlying notional value.

An investor (typically an institution) with exposure to equities in a portfolio is exposed to the risk of the stock market dropping in value. To hedge this risk using equity swaps, the institution could utilize an equity swap promising to pay the (variable) return on a broad stock index such as the S&P 500 in return for a fixed percentage payment based on the notional value of the underlying portfolio of stocks. The resultant cash flows provide a fixed rate of return to the institution that protects the portfolio from falling stock prices. Note, for instance, that falling stock prices generate negative "payments" on the variable component of the swap; that is, the institution's counterparty makes both the fixed payment promised in the swap agreement *and* the payment to compensate for the lost value in the underlying portfolio. Of course, this hedging strategy involves the loss of upside gains from the stock market as well—part of the price paid for hedge protection.

As with most swap agreements, terms of the contract can be tailored to individual needs in the marketplace. Instead of a fixed-for-variable swap, participants might elect to exchange the return on larger stocks, such as the S&P 100 index, for the return on smaller stocks, such as the Russell 2000 index. This strategy could capture the spread between large and small stock returns. Although payments commonly are arranged at quarterly intervals, contract terms can be varied to meet the cash flow objectives of each counterparty.

One permutation of an equity index swap involves a single-stock swap contract wherein one party promises the return on a single firm in exchange for a fixed rate of return. These so-called single-name swaps introduce counterparty credit risk into equity investments (as do most all swap agreements). In the case of single-name swaps, however, this risk presents a significant impediment to the development of this market. For example, during the fall of 2008, when short sales were banned on many individual stocks, market participants could have skirted the ban using equity swaps. Despite the availability of these vehicles, very few took advantage of this chance, largely due to counterparty credit risk considerations.

Single-name swaps also present challenges to the regulatory world that governs corporations in the United States. The SEC regulates stock trading in the United States as the investor's advocate. Single-name swaps, however, obfuscate the true ownership base of a stock. Institutions that own large blocks of the actual stock must report these holdings to the SEC. The counterparty to the single-name swap, however, currently is not required to report to regulatory authorities despite the fact that its position directly benefits from the movement in the underlying stock, just as if it had in fact purchased the stock itself.

Why might this matter? one might ask. Well, the recent case of CSX Industries, a railroad operator, is illustrative. On February 7, 2008, the hedge fund TCI sent CSX a letter stating its intentions to acquire effective control of the railroad operator. Unbeknownst to CSX, TCI and other hedge funds had secretly used equity swaps to effectively take a 12.3 percent ownership stake in the railroad operator (with total direct and indirect ownership of about 20 percent of the company).

CSX subsequently sued these hedge funds, alleging that, by using swap contracts, they effectively evaded filing requirements that govern the disclosure of beneficial ownership of company shares under federal securities law. The SEC disagreed, however, claiming that these swaps were not sufficient to create a beneficial ownership in the firm.

Financial engineering of all types, and swap agreements in particular, present challenges to U.S. legal and regulatory authorities. Securities law, written prior to the conceptual development of swaps and other derivative contracts, is evolving continually to address issues raised by complex derivatives.

The growth of equity swap contracts remains robust. According to statistics provided by the Bank for International Settlements, the total notional value of equity-linked swaps and forwards exceeded $4.0 trillion in June 2008 (BIS 2008a). OTC equity-linked swaps and forward contracts alone totaled more than $2.6 trillion in notional value in June 2008, an increase of 85.8 percent from two years prior. As with most derivatives instruments, OTC equity-linked swaps and forward contracts trade in global markets with Europe taking 51 percent market share, the United States an additional 29 percent, Latin America at 5 percent, and the remainder spread between Japan and other Asian countries. Exchange-traded equity-linked swaps and forwards also exceeded $1.4 trillion in notional value in September 2008.

FUTURE OF EQUITY DERIVATIVES

Assets under management for global hedge funds increased to $2.650 trillion in 2007, according to HedgeFund Intelligence (2007). This represents a rise of 27 percent over the previous year. The Bank for International Settlements indicates that turnover of equity-linked exchange-traded derivatives rose by almost 33 percent in 2007 (BIS 2008b). U.S. equity options trading volume increased almost 50 percent during 2007 as well, according to the Options Clearing Corporation (the U.S.-based clearing corporation (2008). Since the credit crisis of 2008, many investment categories have seen dramatic declines in both price levels and trading volume. Notably, while liquidity in OTC-traded products has fallen drastically during 2008, exchange-traded products have seen record levels of volume and trading activity as large institutions seek the relatively safe haven of central clearing on organized exchanges.

Part of the liquidity problem lies in the design of equity derivative instruments themselves. The nonlinear payoffs that are generated by options contracts, for instance, make it difficult to hedge large portfolios without regular adjustments. These regular adjustments presume adequate liquidity over the holding period of the option, exposing traders to liquidity risk not only at the time of purchase but throughout the life of the option. Further uncertainty exists even if individual assets initially are hedged appropriately because the correlations of assets within a portfolio can and do change over time. As a result, substantial correlation risk can also affect the bottom line for many banks.

The worldwide spread of derivatives contracts has also presented challenges to regulators and legal authorities. The Options Industry Council survey in 2006 indicated that 15 to 20 percent of U.S. options trading volume originated from Europe, creating the need for international cooperation. The CFTC has taken a leadership role in the International Organization of Securities Commissions and has established dozens of memorandums of understanding with other countries to share expertise and data and coordinate on regulatory affairs. The SEC is similarly working on rule changes to make it easier for foreign derivatives exchanges to market their products in the United States.

International cooperation is needed to keep up with the growth and popularity of U.S. equity and equity index options abroad. As electronic access drives information and monitoring costs lower, the world has become a smaller place for derivatives traders. Two major exchanges—the New York Stock Exchange (NYSE)/Euronext and Eurexhave recently invested in new technology that should help to continue driving down costs. NYSE/Euronext, operating options markets in Amsterdam, Paris, London, and New York, is currently attempting to draw more European options markets to join its group. NYSE/Euronext affiliates abroad held an 11 percent market share in U.S. trading during July 2008. By comparison, the Chicago Board Options Exchanges and the electronic International Securities Exchange (ISE, the U.S. subsidiary of Eurex) held approximately 34 percent and 28 percent domestic market shares, respectively.

Indeed, because of an increase in international demand for derivatives there is an increase in pressure for international mergers of exchanges. International mergers at the exchange level have created a dynamic environment for derivatives traders. As competition for trading volume increases, technological improvements facilitate the movement of positions around the globe. The growth of exchange-traded derivatives has increased partly due to credit concerns with counterparty risk in many of the OTC markets. During 2008, there was a shift toward stronger credit risk firms and toward exchange-traded equity derivatives as a direct result of the credit crisis.

REFERENCES

BIS (Bank for International Settlements), 2008a, http://www.bis.org/statistics/otcder/dt22b 22c.pdf.

BIS (Bank for International Settlements), 2008b, http://www.bis.org/statistics/derdetailed.htm.

Burghardt, G. 2008. "Volume Surges Again: Global Futures and Options Trading Rises 28% in 2007," *Futures Industry Magazine*, March/April, 14–26.

Hedge Fund Intelligence, 2007, Global Review Update, Autumn, 6–7.

Nybo, A. 2008. "Rising Institutional Demand for Equity Options," *Futures Industry Magazine*, May/June, 32–34.

Options Clearing Corporation, 2008, http://optionsclearing.com/onn/volume/volume information index.jsp.

Options Industry Council, 2006, http://www.optionseducation.org/about/oic_timeline.jsp." Q&A: Nick Tranter: BNP Paribas Looks to the Future." 2008. Futures and Options World, July 1, 7–8.

FURTHER READING

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance* 56: 1–43.

DeFontnouvelle, P., R. P. H. Fishe, and J. H. Harris. 2005. "How New Entry in Options Markets affected Market Making and Trading Costs," *Journal of Investment Management* 3: 24–40.

ABOUT THE AUTHORS

Jeffrey H. Harris is the chief economist of the U.S. Commodity Futures Trading Commission and professor of finance in the Lerner College of Business and Economics at the University of Delaware. He earned a BA in physics from the University of Iowa in 1986, an MBA from the University of Iowa in 1987, and a PhD in finance from the Ohio State University in 1995. He is the author of 10 published articles in leading peer-reviewed journals on topics related to market microstructure, initial public offerings, and derivatives markets. Dr. Harris has served as Visiting Academic Scholar at the U.S. Securities and Exchange Commission and as Visiting Academic Fellow at Nasdaq. Previously he held faculty appointments at the University of Notre Dame and the Ohio State University. He currently serves on the Program Committee for the Financial Management Association International, the Western Finance Association and the European Finance Association.

L. Mick Swartz graduated with a PhD in finance from the University of Iowa in 1994. That year he taught for the University of Manitoba in Winnipeg, Canada, and taught finance in Ukraine for the Canadian government. Dr. Swartz has received many teaching awards for teaching corporate finance, investments, and derivatives (options and futures). He has over 10 refereed journal articles on corporate finance/governance and investments and over 20 presentations at finance conferences. He has been a speaker for community groups, consulted state legislators concerning executive compensation, given advice on merger legislation, and taught in Australia, the Ukraine, the United Kingdom, Canada, Poland, and Germany. Dr. Swartz has appeared on NBC and CNN as well as many radio broadcasts, discussing topics such as the financial crisis, executive pay, investment strategies, retirement planning, bailouts, derivatives, options, SWAPS, and the general economic environment.

CHAPTER 8

Foreign Exchange Derivatives

ROBERT W. KOLB
Professor of Finance and Frank W. Considine Chair of Applied Ethics,
Loyola University Chicago

By virtually any measure, the foreign exchange market is enormous. In 2007, the daily turnover in foreign exchange markets averaged \$3.2 trillion daily. This figure includes spot transactions in which one currency is exchanged for another for immediate delivery, and the spot exchange market accounts for slightly more than \$1 trillion of this total (Bank for International Settlements [BIS], 2007, p. 4). But it also includes activity in foreign exchange (FX) derivatives markets. These derivative markets include forwards, futures, options, swaps, and a variety of other more esoteric instruments. This chapter briefly surveys the pricing principles for foreign exchange contracts and discusses the principal types of foreign exchange derivatives.

BASIC PRICING PRINCIPLES

Like many other financial instruments, the markets for foreign exchange have very few frictions and are characterized by a very high degree of transparency and low transaction costs. As a result, pricing of foreign exchange instruments in actual practice correspond closely to theoretical values, although there are some significant departures. The two most important pricing principles in foreign exchange are the purchasing power parity theorem and the interest rate parity theorem. Although these principles have distinct names, they both specify no-arbitrage conditions, and they have close analogues in other markets.

Purchasing Power Parity Theorem

The purchasing power parity theorem (PPPT) essentially claims that prices for identical goods trading in different countries with different currencies must have the same cost, except for factors due to market frictions. If this condition were not met, there would be arbitrage opportunities. For example, if a widget in England sold for a higher price, considering exchange rates, than a widget in the United

---

This chapter draws on Robert W. Kolb and James A. Overdahl, *Futures, Options and Swaps*, 5th ed. (London: Blackwell Publishers, 2007).

States, a trader could acquire the widget in England by paying pounds, transport the widget to the United States, sell it for dollars, and convert the dollars to more pounds than the widget cost in England.

Thus, the price of a widget in the United States and England must be identical, except for market frictions. In our widget example, one clear kind of market friction is the cost of transporting the widget across the Atlantic. Other important kinds of transaction costs that arise in international trade would include tariffs, taxes, import quotas, and the like. As a result, in actual practice, the same goods can trade for substantially different prices in different countries. But the possibility of arbitrage keeps those pricing discrepancies to a level that reflects only market frictions.

The *Economist* magazine frequently publishes its Big Mac Index, which compares the prices of McDonald's Big Mac sandwich in different currencies. In the summer of 2008, the U.S. price of a Big Mac was \$3.57. At the same time, the equivalent price in China was \$1.83, \$4.57 in England, and \$7.88 in Norway. These prices obviously represent enormous departures from the PPPT. With the very real difficulties in conducting arbitrage for Big Macs between countries, there are few market forces that can erase these huge differentials. By contrast, for goods that are easy to transport, that have less significant market frictions, those differentials between countries will be much lower and there will tend to be a single world price for such goods. Thus, for most commodities, there tends to be a single world price, yet this can be disturbed by transportation costs and government-imposed costs that frustrate the PPPT.

Interest Rate Parity Theorem

The interest rate parity theorem (IRPT) pertains to the differentials between spot exchange rates and forward exchange rates. The spot exchange rate is the rate of exchange between two currencies for immediate delivery, while the forward exchange rate is the rate at which one can contract today for the exchange of currencies at a specified future date. The IRPT asserts that the spot exchange rate and the forward exchange rate between two currencies, and the interest rates on the two currencies form a system of prices that must bear a certain relationship to one another to prevent arbitrage.

The theorem can be understood best through an example. Assume that the spot exchange rate is \$1.42 = €1, that the forward rate of exchange for delivery in one year is \$1.35 = €1, that the one-year interest rate for dollars is 5 percent and the one-year interest rate on the euro is 6 percent. Faced with these exchange rates and interest rates, a European investor would borrow with €1 at the 6 percent interest rate, exchange the euro for \$1.42, and invest the dollar proceeds in the United States at the 5 percent rate for one year, giving anticipated dollar proceeds in one year of \$1.4910. Simultaneously, the investor would sell \$1.4910 in the forward market at the one-year forward rate of €1 = \$1.35 for total proceeds of €1.1044. From this sum, the investor would repay the euro that she borrowed for €1.06, leaving a total profit of €0.0444.

This result is clearly an arbitrage profit, which indicates a violation of IRPT. Thus, the system of exchange rates and interest rates constitutes an inconsistency that can be resolved by the adjustment of any of the four terms. Barring market frictions, the activity of arbitrageurs would generate market forces that would restore the system of exchange rates and interest rates to conformity with the IRPT.

As we have seen, the PPPT does not necessarily hold very well in actual practice. In contrast, IRPT holds extremely well, and FX prices correspond extremely closely to the conditions specified by the IRPT.

FOREIGN EXCHANGE FORWARD AND FUTURES CONTRACTS

The foreign exchange forward market is an over-the-counter (OTC) market that trades on a worldwide basis and is dominated by large financial institutions. Typical contract amounts are quite large, and the market sees very few individual investors. By contrast, there are foreign exchange futures markets in many countries. In comparison with the forward market, these futures markets are quite small. The discussion of these markets will focus on the forward market, which will be contrasted with a single foreign exchange futures market, the Chicago Mercantile Exchange in the United States.[1]

The Bank for International Settlements (BIS 2007, p. 4) reports estimates of market size in its surveys of central banks. For 2007, the BIS reported that the average daily turnover for foreign exchange forwards was $362 billion. Exhibit 8.1 shows the main currency pairings in the foreign exchange forward market. In spite of the emergence of the euro as an important international currency, the U.S. dollar continues to dominate the market, with fully 71 percent of all foreign exchange transactions involving the dollar as one of the currencies in each pair. As Exhibit 8.1 shows, the dollar/euro currency pair is the most heavily traded type of transaction in the FX forward market. Contract maturities in the OTC FX forward market are determined by agreement between the contracting parties and can be of any duration. Typical terms are 90, 180, or 270 days and one year, and the normal contract size is reckoned in the millions to many millions of dollars. Most FX forward contracts are settled by actual delivery.

The Futures Industry Association (2008) reported that worldwide volume in foreign exchange futures in 2007 was 335 million contracts. With total futures contract volume across the world and all kinds of instruments of 15.18 billion, foreign exchange futures account for only 2.2 percent of the total. As mentioned at the outset of our discussion, we take the foreign exchange futures market of the Chicago Mercantile Exchange (CME) as an example. The CME trades a wide

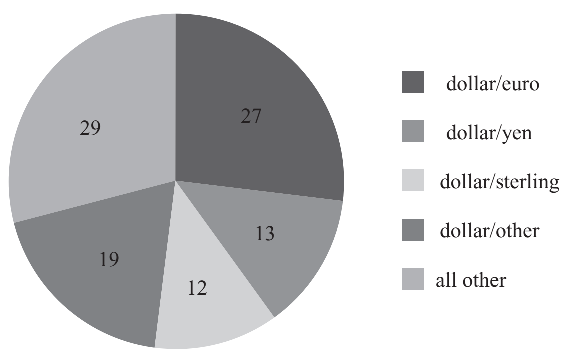


Exhibit 8.1 Foreign Exchange Forwards by Currency Pairs (%)

variety of futures contracts. Compared to the OTC market for forwards, all of the futures volumes are quite small. The main volume is concentrated in the currencies of major industrialized countries: the United States, European Union, Switzerland, United Kingdom, and Japan. However, the CME lists futures for many relatively obscure currency pairs, such as the dollar/ruble and the dollar/shekel. In general, these markets see little volume and are relatively illiquid.

In contrast with the OTC FX market, the CME FX futures trade with specified contract maturities, usually on the March/June/September/December cycle, with each contract having a fixed quantity. Contract sizes are relatively small, with $100,000 being a representative figure for the underlying currency value of each contract. As with all futures, the CME imposes margin requirements, and frequent cash settlements to avoid accumulating credit exposure. In contrast with OTC FX forwards, most FX futures are settled by offset rather than delivery.

For both FX forwards and futures, pricing generally follows the strictures of the IRPT very closely. FX forward and futures prices for the same currency pairs and maturities are virtually identical, with almost all of the small difference being attributable to the different institutional features of the markets, notably the requirement for margins and daily resettlement in the futures market but not in the OTC market.

As is typical with almost all futures and forwards, FX futures and forwards serve as vehicles for both speculation and hedging. Speculative motives turn on anticipation of changes in interest rates, inflation rates, central bank interventions, and speculative attacks on currencies with fixed or pegged exchange rates. Hedging motivations are usually classified as being oriented toward either *transaction exposure* or *translation exposure.* Transaction exposure refers to an anticipated future exchange of currencies. For example, an importer might expect to pay for a shipment of goods in a foreign currency at a future date. The importer might hedge the currency risk associated with this future transaction by using the forward or futures market. Translation exposure arises when transactions denominated in a foreign currency must be translated, or restated, in the local currency. Thus, translation exposure is an accounting vulnerability. Yet some firms enter the FX futures or forward markets, thereby engaging in a transaction of economic significance, to avoid the risk of having reported earnings adversely affected by a disadvantageous currency fluctuation.

FOREIGN EXCHANGE OPTIONS

Much of the preceding discussion of FX forwards and futures also apply to FX options. The OTC FX options market dwarfs exchange-traded options. The OTC FX options market is a worldwide market dominated by institutional investors, with contract maturities and quantities being flexible and determined by negotiation. BIS (2007, p. 14) gauges daily average turnover of FX options on the OTC as $212 billion in 2007. As with forwards and futures, the main markets for FX options involve the currencies of the large industrial countries. However, virtually any kind of option with any pair of currencies is available in the OTC market upon request. Exchanges generally trade FX options on all of the currency pairs for which they list FX futures. The common pricing relationships between options and futures stimulate traders to trade FX options and futures in concert. While most of the trading volume of FX options is concentrated in plain vanilla options, there is a robust market in exotic FX options as well. As with forwards and futures, the OTC market for foreign exchange dwarfs trading on organized exchanges, with average daily turnover in OTC-traded FX options exceeding \$31 billion in 2007 (BIS 2007, p. 15).

FX OPTION PRICING

For plain vanilla FX options, pricing generally conforms quite closely to appropriately adjusted Black-Scholes model values, and models of this type are used throughout the market to price FX options. While the Black-Scholes options pricing model strictly pertains only to options on stocks that pay no dividends, FX option models using the Black-Scholes technology developed quickly following the publication of the original Black-Scholes paper.[2]

The essential insight was to see that the existence of a continuous payoff from an asset, whether it is a continuous dividend rate or a rate of interest, was to treat that outflow as a diminution of the interest rate in the original Black-Scholes model.

$$
\begin{aligned}
C_t &= e^{-r_F(T-t)} FC\, N(d_1) - X e^{-r_d(T-t)}\, N(d_2) \\
P_t &= X e^{-r_F(T-t)}\, N(-d_2) - FC\, e^{-r_F(T-t)} N(-d_1) \\
d_1 &= \frac{\ln(FC/X)\,(r_D - r_F + 0.5\sigma^2)\,(T-t)}{\sigma\sqrt{T-t}} \\
d_2 &= d_1 - \sigma\sqrt{T-t}
\end{aligned}
$$

where
C_t = price of a call option (priced in the domestic currency) on foreign currency FC
P_t = price of a put option (priced in the domestic currency) on foreign currency FC
FC = a quantity of the foreign currency
r_D, r_F = domestic and foreign interest rates, respectively
X = exercise price
$T - t$ = time until expiration
σ^2 = variance of the foreign currency value
$N(\bullet)$ = cumulative normal function

The essential difference between this equation and the pricing of a stock paying a continuous dividend rate is that the interest rate on the foreign currency, r_F, takes the place of the dividend rate. This extension of the original Black-Scholes model functions extremely well for FX options, and this model is an industry standard. In addition to the model of the equation, binary methods also function extremely well and are widely used for pricing more complex options.

PLAIN VANILLA FOREIGN EXCHANGE SWAPS

There are two essential types of plain vanilla swap agreements: interest rate swaps and foreign exchange swaps. In a plain vanilla interest rate swap, one party pays a floating rate of interest, while the other pays a fixed rate of interest on the same quantity of a currency (the nominal amount). Further, at each payment date, the obligations of the two parties typically are netted out, and only the netted amount actually changes hands. In a plain vanilla interest rate swap, these periodic payments are the only cash flows; the nominal amount does not change hands. In a plain vanilla foreign exchange swap, there are always two currencies involved, and the parties generally exchange the foreign currencies that constitute the nominal amounts as well. In contrast to interest rate swaps, plain vanilla foreign exchange swaps often are motivated by the desire of the parties actually to acquire the other currency, and the full amounts of the periodic payments are made by both parties in the respective currencies. This market is extremely robust, with average daily turnover of nominal amounts exceeding $2.2 trillion in 2007 (BIS 2007, p. 85).

Today, there are no officially imposed restrictions on the movement of most major currencies. In the not too recent past, however, central banks in many industrialized countries imposed active restrictions on the flow of currency. The parallel loan market developed to circumvent restrictions imposed by the Bank of England on the free flow of British pounds. British firms wishing to invest abroad generally needed to convert pounds into U.S. dollars. The Bank of England required these firms to buy dollars at an exchange rate above the market price. The purpose of this policy was to defend the value of the pound in terms of other currencies. Firms, naturally, were not interested in subsidizing the Bank of England by paying the above-market rate for dollars required by the Bank of England's policies. Attempts to evade these currency controls led directly to the development of the market for currency swaps.

In this environment of foreign exchange controls, consider two similar firms, one British and one American, each with operating subsidiaries in both countries. By cooperating with a U.S. firm that has operations in England, the British firm can evade the currency controls. The British firm lends pounds to the U.S. subsidiary operating in England, while the U.S. firm lends a similar amount to the British subsidiary operating in the United States. This is a parallel loan—two multinational firms lend each other equivalent amounts of two different currencies on equivalent terms in two countries. A parallel loan is also known as a back-to-back loan.

The development of swaps stemmed directly from these incentives to create parallel loans. Although the projected cash flows from the parallel loan and the plain vanilla currency swap are identical, there are still some subtle but important differences. In the parallel loan, both parties need to pretend that the transactions are completely distinct. If so, default on one of the loans would not justify default by the other party. In an interest rate swap agreement, there are cross-default clauses. (The parallel loan cannot contain those cross-default clauses, because the two parties are trying to pretend that the parallel loans are distinct.) Also, a swap agreement would have lower transaction costs than arranging two separate loans.

In a plain vanilla FX swap, the parties exchange currencies at the outset of the swap's tenor and pay interest on the currency received (which can be at either a fixed or floating rate). At the termination of the swap, the parties again exchange the identical nominal currency amounts. Considering fixed and floating rates, there are four possible basic payment patterns which can be illustrated by reference to two parties, A and B, and an FX swap between dollars and euros:

1. Party A pays a fixed rate on dollars received, and Party B pays a fixed rate on euros received.
2. Party A pays a floating rate on dollars received, and Party B pays a fixed rate on euros received.
3. Party A pays a fixed rate on dollars received, and Party B pays a floating rate on euros received.
4. Party A pays a floating rate on dollars received, and Party B pays a floating rate on euros received.

Although all four patterns of interest payments are observed in the market, the predominant quotation is of the second type: pay floating on dollars/pay fixed on the foreign currency. This is known as the plain vanilla currency swap. Of these, the simplest kind of currency swap arises when each party pays a fixed rate of interest on the currency it receives, such as type 1. The fixed-for-fixed currency swap involves three different sets of cash flows. First, at the initiation of the swap, the two parties actually exchange cash. Typically, the motivation for the currency swap is the actual need for funds denominated in a different currency. This differs from the interest rate swap in which both parties deal in a single currency and can pay the net amount. Second, the parties make periodic interest payments to each other during the life of the swap agreement, and these payments are made in full without netting. Third, at the termination of the swap, the parties again exchange the principal.

Pricing of FX swaps must reflect the term structure of interest rates for both currencies as well as the term structure of FX spot and forward rates. These relationships are mediated by the interest rate parity theorem discussed earlier. Fair value FX swaps leave the two parties with no anticipated change in wealth given these pricing relationships. Of course, deviations in interest and exchange rates from their respective forward rates will determine the actual wealth outcomes for the swap.

FLAVORED CURRENCY SWAPS

As we have noted, a plain vanilla currency swap calls for the two counterparties to exchange currencies at the outset of the swap and to make a series of interest payments for the currency that is received. However, currency swaps are subject to many elaborations. More complicated swap structures can be created by allowing the notional principal to vary over the tenor of the swap. For example, currency swaps can be amortizing (the notional amount decreases over the swap's tenor), accreting (the notional amount increase over the swap's tenor), seasonal (the notional amount fluctuates over the calendar year in a specified pattern), and roller-coaster swaps (the notional amount varies based on prenegotiated changes in the notional principal over the tenor of the swap).

For example, a U.S. firm might import clothes from Hong Kong, with much higher imports in the winter, and pay for these in Hong Kong dollars. The firm might structure a currency swap with a seasonal notional principal to match its greater anticipated need for funds in winter months. Two fixed-for-floating swaps can be combined to create a fixed-for-fixed currency swap. As a second example, a CIRCUS swap is a fixed-for-fixed currency swap created by combining a plain vanilla interest rate swap with a plain vanilla currency swap. (CIRCUS stands for *c*ombined *i*nterest *r*ate and *cu*rrency *s*wap.)

A dual-currency bond has principal payments denominated in one currency, with coupon payments denominated in a second currency. For example, an issuing firm might borrow dollars and pay coupon payments on the instrument in euros. When the bond expires, the firm would repay its principal obligation in dollars. This dual-currency bond can be synthesized from a regular single-currency bond with all payments in dollars (a dollar-pay bond) combined with a fixed-for-fixed currency swap.

A currency annuity swap is similar to a plain vanilla currency swap without the exchange of principal at the initiation or the termination of the swap. It is also known as a currency basis swap. For example, one party might make a sequence of payments based on British London Interbank Offered Rate (LIBOR) while the other makes a sequence of payments based on U.S. LIBOR. As we will see, the currency annuity swap generally requires one party to pay an additional spread to the other or to make an up-front payment at the time of the swap. Variations of this structure can be created by allowing one, or both, parties to pay at a fixed rate. In pricing these swaps, the key is to specify a spread or up-front payment that makes the present value of the cash flows incurred by each party equal.

In a cross-index basis note, or quanto note, the investor receives a rate of interest that is based on a floating rate index for a foreign short-term rate but is paid in the investor's domestic currency. For example, a U.S. investor might buy a note with an interest rate based on European interest rates, but with all payments on the note being made in U.S. dollars. From the investor's point of view, the quanto note allows exposure to foreign interest rates without currency exposure. Also, the quanto note allows the investor to speculate on relative changes in the foreign and domestic yield curves. From the point of view of the issuer, the quanto note can offer investors attractive investment opportunities that might not be available elsewhere. Also, the issuer can issue a quanto note but use swaps to transform its risk exposure to a perhaps more congenial form.

CONCLUSION

This chapter has provided a brief overview of the range of foreign exchange derivatives current in the market today. The basic no-arbitrage pricing principles common to all financial derivatives apply with full force to FX derivatives. As we noted, the Black-Scholes model, adapted for the two interest rates involved in an FX option, performs extremely well as a practical method for pricing FX options. As we have seen, the full range of derivative types is available for foreign exchange as one asset class among many. These range from forwards to futures, to options and swaps. For example, there is a robust market for exotic options based on foreign exchange as well, with active volume for these instruments in the OTC market.

ENDNOTES

1. Even though the strong dominance of OTC trading continues in the FX market, there is a movement toward more exchange-based trading, especially on electronic platforms. See Acworth (2007).
2. The essential advancement over the original Black-Scholes article was made by Robert C. Merton (1973), and this consisted of extending the model to account for continuous dividends. Specific application of this framework to currency options was made by Garman and Kohlhagen (1983).

REFERENCES

Acworth, W. 2007. "Foreign Exchange Trading: New Trading Platforms Reshape FX Markets," *Futures Industry* (September–October): 20–23.

Bank for International Settlements. 2007. *Triennial Central Bank Survey: Foreign Exchange and Derivatives Market Activity in 2007* (December).

DeRosa, D. F. 2000. *Options on Foreign Exchange*. New York: John Wiley & Sons.

Futures Industry Association. 2008. *Futures Industry* (March/April): 16.

Garman, M. B., and S. V. Kohlhagen. 1983. "Foreign Currency Option Values," *Journal of International Money and Finance* (December): 231–237.

Kolb, R. W., and J. A. Overdahl. 2007. *Futures, Options and Swaps*, 5th ed. Oxford: Blackwell Publishers.

Lipton, A. 2001. *Mathematical Methods for Foreign Exchange: A Financial Engineer's Approach.* Hoboken, NJ: John Wiley & Sons.

Merton, R. C. 1973. "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* (Spring): 141–183.

Weithers, T. 2006. *Foreign Exchange: A Practical Guide to the FX Markets*, Hoboken, NJ: John Wiley & Sons.

ABOUT THE AUTHOR

Robert W. Kolb is professor of finance at Loyola University Chicago, where he also holds the Frank W. Considine Chair of Applied Ethics. Dr. Kolb received two PhDs from the University of North Carolina at Chapel Hill in philosophy and finance, (philosophy 1974, finance 1978). Previously he held full-time academic positions at the University of Florida, Emory University, the University of Miami, and the University of Colorado. Over his career, he has published more than 50 academic research articles and more than 20 books, most focusing on financial derivatives and their applications to risk management. In 1990, he founded Kolb Publishing Company to publish finance and economics university texts. He sold the firm to Blackwell Publishers of Oxford, England, in 1995.

CHAPTER 9

Energy Derivatives

CRAIG PIRRONG
Professor of Finance, Director, Global Energy Management Institute,
Bauer College of Business, University of Houston

INTRODUCTION

Energy derivatives are a relatively new financial product but have been growing rapidly since their introduction in the 1970s. Large quantities of energy derivatives are traded on exchanges in the over-the-counter (OTC) market. Moreover, in the 2000s, there has been a notable influx of energy trading by financial institutions and portfolio investors that increasingly view commodities in general, and energy in particular, as a distinct asset class that exhibits a low (and indeed, negative) correlation with traditional equity and fixed income investments, and therefore improves portfolio risk-return characteristics through the effects of diversification. Commodities are also viewed as an inflation hedge. This development, which has occurred in parallel with a fourfold increase in oil prices since 2003, has sparked intense criticism of energy derivatives trading.

PRODUCTS: AN OVERVIEW

Energy derivatives are traded on a variety of energy products. The largest single category consists of derivatives on petroleum and petroleum derivatives. Petroleum derivatives include contracts on crude oil, with sweet crudes (West Texas Intermediate and Brent) being by far the dominant products; middle distillates, including heating oil, gasoil, and jet fuel; and gasoline. In addition, derivatives are traded on natural gas and various gas liquids (such as propane) and electricity. Although not properly an energy product, derivatives on emissions (such as sulfur dioxide and carbon dioxide) are also traded; since these emissions are often the result of energy production or consumption (e.g., the generation of electricity produces SO_2 and CO_2), emissions derivatives are often bought and sold by energy trading desks.

All types of standard derivatives are traded on energy products. Standardized futures contracts are traded on exchanges, most notably the New York Mercantile Exchange and ICE Futures, a division of the Intercontinental Exchange (ICE). Energy futures contracts have been introduced by the Shanghai Futures Exchange and the Dubai Mercantile Exchange, and the Hong Kong Futures Exchange and the Singapore Mercantile Exchange have announced efforts to introduce energy futures contracts.

In addition, options on futures contracts are traded on these exchanges. Furthermore, there is extensive trading in OTC energy derivatives instruments. These include vanilla forward contracts and swaps, basis and differential swaps, and options and swaptions. These are primarily cash settled instruments, with final settlement prices based on exchange settlement prices or prices reported in industry publications.

Some energy products, such as natural gas, are costly to transport relative to value, and hence for these products there are often numerous local markets. So-called basis contracts have developed to facilitate hedging and price discovery for these local markets. These contracts have payoffs that are equal to the difference between prices in two different markets. The most common example of such contracts have payoffs equal to the difference between the price of natural gas at the Henry Hub (the delivery point for New York Mercantile Exchange [NYMEX] natural gas futures) and the price in another market (such as the Chicago City Gate). Settlement prices in these contracts are typically based on indices reported in trade publications such as *Inside FERC* or *Natural Gas Daily.* These publications collect and average prices for cash transactions; *Inside FERC* reports prices on contracts for delivery of gas ratably over a contract month entered into by buyers and sellers during the "bid week" prior to the delivery month, whereas *Natural Gas Daily* reports cash prices on transactions for delivery the following day.

HISTORY

Futures contracts were traded on oil in the United States during the nineteenth century, especially in the aftermath of the first Pennsylvania oil boom. Numerous oil futures exchanges came and went during this period (Wiener 1992). The development of the Standard Oil monopoly and the pervasive move to vertical integration in the industry undermined the need for futures trading. Energy derivatives trading largely disappeared in the late nineteenth century and remained largely moribund until the 1970s (although there were attempts to trade oil futures in the 1930s).

The New York Cotton Exchange (NYCE) introduced first modern exchange traded energy contract, on propane, in 1971. The NYMEX introduced the first widely traded energy futures contract (on heating oil) in 1974.

Developments during the 1970s, notably the decade's oil shocks and the abrogation of long-term "equity" oil contracts between producing nations and major oil companies, disrupted traditional oil marketing methods and encouraged the development of an oil spot market. This, in turn, created a need for energy derivatives contracts to facilitate risk management and price discovery. In short order, exchanges introduced a variety of futures contracts on petroleum products. NYCE introduced a crude oil contract in 1974, and the Chicago Board of Trade and the NYMEX launched crude contracts in 1983. NYMEX soon gained the lead in this market. Although other exchanges launched oil contracts in the 1980s, NYMEX maintained its dominance, and extended it with the introduction of gasoline futures in 1984.

Outside the United States, the United Kingdom's International Petroleum Exchange (IPE, since acquired by ICE and now operating as ICE Futures) began trading gasoil futures in 1980 and introduced its Brent crude oil futures contract in 1988.

Deregulation of transportation and prices in the United States during the 1980s laid the foundation for natural gas futures trading. NYMEX launched the successful Henry Hub natural gas futures contract in 1990. IPE launched a natural gas futures contract for delivery into the U.K. grid in 1997.

Deregulation was also the impetus for the introduction of futures contracts on electricity. NYMEX began trading California-based electricity futures contracts in 1996 (when the state was in the process of a major regulatory restructuring) and added contracts for midwestern and eastern markets two years later. IPE introduced UK electricity futures in 2004, and on the continent, the European Electricity Exchange commenced trading futures in 2001. Nordpool began trading futures on Scandinavian electricity in the 1990s, and the Amsterdam Power Exchange was created in 1999.

Development of OTC trading in the energy complex paralleled that of exchange trading, with crude oil and refined products trading developing first, followed by natural gas and then electricity.

PETROLEUM DERIVATIVES: DETAILS

There are two dominant crude oil futures contracts, West Texas Intermediate (WTI) traded on NYMEX and Brent traded on ICE Futures (the successor to the IPE). Both contracts are for light, sweet (i.e., low-sulfur) crude oils. Such crudes are the most valuable but represent a smaller and smaller fraction of world crude oil production.

The WTI contract is a delivery settled contract, calling for delivery of 1,000 barrels of crude oil in Cushing, Oklahoma (a major storage and transportation hub). ICE Futures also trades a cash-settled WTI contract, with contract settlement price tied to the NYMEX price. Brent is also delivery settled, but with a cash settlement option, with the cash settlement index based on transactions prices for 21-day Brent, Forties, Osberg, and Ekofisk (BFOE) cargoes.

Trading volume in these contracts has grown dramatically in recent years. For instance, NYMEX crude oil average daily volume was 107,579 contracts in 1988, 254,162 contracts in 1998, and 1,173,458 by June 2008.

WTI and Brent futures are the primary beacons of price discovery in the oil market. Many physical oil trades are quoted basis one of these futures instruments, with physical transactions executed at differentials to the futures prices. Since sweet crudes are becoming progressively less representative of world oil production, there is a perceived need for a pricing benchmark based on a heavier, sourer crude. Despite several attempts to introduce such contracts, WTI and Brent have retained their dominance due to the difficulty of shifting liquidity to new contracts.

NYMEX's most important refined product futures—for heating oil and gasoline—are settled by physical delivery of 42,000 gallons of product in New York Harbor (a major refining center). Gasoline futures contract specifications have changed throughout the years due to changes in American environmental regulations. ICE Futures gasoil is settled by delivery in barges in Amsterdam, Antwerp, and Rotterdam.

Spread trading in petroleum futures is quite common. For instance, a "crack spread" trade involves the simultaneous purchase of a crude oil futures contract and the sale of gasoline and/or heating oil futures contracts. Such trades can be used to hedge or speculate on refining margins.

Both NYMEX and IPE were originally floor-based exchanges. IPE switched all trading to an electronic platform in 2005, and NYMEX entered into an agreement with the Chicago Mercantile Exchange to trade NYMEX petroleum futures contracts on the GLOBEX system in 2006. Although NYMEX continues to trade crude and refined products on its New York floor, over 80 percent of volume in these products is traded electronically on GLOBEX.

Large volumes of crude oil forwards, swaps, and options are traded in the OTC market. Most swaps and forwards are cash settled, with NYMEX WTI futures prices being an important source of settlement prices for these contracts. In addition, there is extensive OTC paper trading of Brent crude or, more recently, BFOE contracts. Twenty-one-day BFOE contracts call for delivery of crude oil in a 21-day window of a particular month. Once the loading date for a particular cargo is set, the contract becomes a "dated BFOE" contract. Like most exchange traded futures contracts, 21-day Brent contracts are liquidated—"rung out"—prior to delivery. In addition, contracts for differences (CDFs), with payoffs based on the difference between dated BFOE and 21-day BFOE, are widely traded; these CFDs are essentially swaps.

NATURAL GAS DERIVATIVES: DETAILS

The main NYMEX natural gas futures contract is settled by delivery of 10,000 million British Thermal Units (MMBTU) of natural gas ratably throughout the contract month at the Henry Hub in Louisiana. NYMEX natural gas basis contracts are settled financially based on the difference between a price index quoted for a particular location and the final settlement price of the NYMEX Henry Hub natural gas contract. For example, the Chicago City Gate settlement is set equal to *Natural Gas Intelligence* (an industry publication) Chicago City Gate index price minus the final settlement price of the Exchange Henry Hub natural gas futures contract for the corresponding month on the last trading day.

NYMEX also trades index swaps, which have payoffs equal to the difference between the average of a daily price during a particular month at a particular location and a monthly forward price at that location. For example, the Chicago City Gate Index Swap settlement is set equal to the average of the Platts *Gas Daily* price at the Chicago City Gate during the contract month minus the bidweek price provided by *Natural Gas Intelligence* Chicago City Gate index price. ("Bidweek" is a period during which buyers and sellers contract for physical natural gas to be delivered in the following calendar month.)

The ICE Futures natural gas contract is also delivery settled, with shorts making delivery ratably throughout the contract month into the British national pipeline grid.

NYMEX Henry Hub natural gas contracts are traded both on the floor and on the GLOBEX system, with the bulk of the volume traded on the latter. The basis swap and index swap contracts are traded on the NYMEX ClearPort system or on the OTC market and cleared through NYMEX. ICE Futures natural gas contracts are traded electronically.

Natural gas trading volumes have increased steadily since the introduction of the Henry Hub contract in 1990. In 1991, the average daily trading volume for this contract was 1,654 lots. By 2001, this volume had grown to 47,457 lots per day. By mid-2008, trading volume averaged 153,329 contracts each day.

Henry Hub swaps, basis swaps, index swaps, and natural gas options and swaptions are also traded extensively in the OTC market. These contracts are traded via voice brokers and on the Intercontinental Exchange's electronic system.

ELECTRICITY DERIVATIVES: DETAILS

The first generation of U.S. electricity futures contracts were settled by delivery of a fixed quantity of electricity in each "peak" hour during the delivery month. These contracts never garnered significant trading volumes, and eventually became dormant.

In the late 1990s and early 2000s, as part of the ongoing industry restructuring process, major Independent System Operators (ISOs) established centralized real-time markets for electricity in large parts of the United States. These ISOs include PJM (originally operating a market in Pennsylvania-New Jersey and Maryland, but now extended to encompass most of the Midwest as well), NYISO (New York), and NEISO (New England). These ISO collect bids from generators indicating the prices at which they are willing to produce power and use these bids to determine the marginal price to generate the electricity demanded in the ISO control region.

New NYMEX futures contracts utilize these "real time" prices established by the ISOs to settle futures contracts. For instance, the final settlement price for the NYMEX PJM peak futures contract for a particular contract month is the average of real time prices at 111 pricing points in the western part of the PJM region during the peak hours of that month.

Volumes of exchange-traded electricity futures contracts have been modest, which reflects the nature of electricity. Although electricity spot prices are highly volatile, electricity forward prices are much less so because (a) the lack of storability attenuates the connections between spot and forward prices, and (b) most of the important shocks that drive electricity spot prices (e.g., weather) are highly transitory. Moreover, evidence suggests that (a) there are substantial hedging imbalances, with long hedgers predominating and (b) it is difficult for speculators to trade electricity (Bessembinder-Lemmon 2002; Pirrong 2008a).

Electricity swaps (with the floating leg of the payoff typically based on ISO spot prices) are traded via voice brokers and on the ICE electronic platform.

PRICING

Energy derivatives exhibit distinctive pricing behaviors. Moreover, standard cost-of-carry models do a poor job of characterizing energy forward curves. This is almost self-evident for electricity; because it cannot be stored, "carry" is infinitely costly, and the cost-of-carry model is therefore inapplicable. Even for ostensibly storable energy commodities, however, the cost of carry model performs badly.

Crude oil is an exhaustible resource, and according to the standard pricing theory (Hotelling, 1931) its price should rise at the rate of interest over time. This, in turn, should lead to contango in the forward market (with deferred futures prices exceeding nearby prices). However, crude oil has historically been in backwardation (with nearby prices exceeding deferred prices) about 60 percent of the trading days from 1983 to 2007. The main periods of contango have occurred when prices were very low, as during the Asian economic crisis of 1998. The one exception to this is the period from late 2005 to 2007, when the crude forward curve was in contango even when prices were at (then) historically high levels.

Various theories attempt to explain the prevalence of backwardation in oil futures. Litzenberger and Rabinowitz (1995) argue that since the owner of an oil well has the real option to determine the timing of extraction, price uncertainty generates option time value that induces the owner to defer production if the futures price is above the spot price by the cost of carry; the forward premium compensates for the deferral of the receipt of revenue from sale, and the owner retains the option to sell at more favorable prices in the future. The only way to induce current production, according to Litzenberger and Rabinowitz, is for the market to trade below full carry and perhaps in backwardation. The deviation from full carry must be sufficient to offset the effect of the option value associated with deferring production. Litzenberger and Rabinowitz document that, consistent with this theory, deviations from full carry are larger when output is high and when volatility is high.

Other explanations of crude oil forward price curves include applications of the standard convenience yield version of the theory of storage (Pindyck 1994), and modifications of the dynamic programming version of the theory of storage (Pirrong 2008b).

In the convenience yield version, inventories of a commodity such as crude generate a stream of benefits accruing to the inventory holder. This functions as an implicit dividend that the owner of the spot captures but the holder of the futures does not, which depresses the futures price relative to the spot price.

In the dynamic programming version, storage is used to smooth out the effects of demand and supply shocks; inventories rise when demand is unexpectedly low (or supply is unexpectedly high) and fall when the reverse is true. Forward curves adjust to provide the incentive for agents to do the right thing; the market moves toward full carry in low-demand/high-supply states in order to reward storage and shifts into backwardation in high-demand/low-supply states to punish holders of inventory, thereby inducing them to liquidate their holdings. In the standard variant of this theory, the volatilities of demand and supply shocks are constant. This variant of the model cannot readily account for phenomenon like that observed in 2005 to 2007, when both inventories and prices were high. Pirrong (2008b) modifies the standard model to allow for stochastic fundamental volatility. This modified model can account for simultaneously high prices and inventories. Positive shocks to fundamental volatility induce agents to hold additional precautionary inventories. For inventories to increase, prices must rise to reduce consumption and increase output.

Other storable energy commodities, notably heating oil, gasoline, and natural gas, exhibit seasonality in their forward curves. This, in turn, reflects the highly seasonal demand in the demand for these commodities.

Electricity forward prices also exhibit substantial seasonality (Eydeland and Wolniac, 2004). The seasonal peaks depend on the markets. In the U.S. South, Midwest, and West, for instance, forward prices are highest for summer delivery months (July and August), lowest in the "shoulder" months (April and October), and exhibit a smaller seasonal peak in winter months (notably December and

January). In contrast, in European (especially Nordic) markets and in the U.S. Northeast, the largest peaks correspond to the winter months.

Energy commodities also exhibit rich relationships. For instance, the crack spread is a measure of the profitability of refining crude oil into the products individuals actually consume to power vehicles or heat homes. Thus, the relationship between crude oil, heating oil, and gasoline prices varies to reflect the economics of transforming the input (crude) into the output refined products. As an example, a cold snap tends to increase the demand for heating oil, the demand for refining capacity, and, indirectly, the demand for crude oil (the demand for crude being derived from the demand for refined products). Thus, this weather shock tends to cause the price of heating oil to rise, the price of crude to rise, and the crack spread to increase (due to the increase in the demand for refining capacity). That is, the price of refined products rises relative to the demand for crude oil. Interestingly, the shock to heating oil demand tends to cause the gasoline price to fall because due to the fact that heating oil and gasoline must be produced in (roughly) fixed proportions, increasing heating oil output also increases gasoline output, which has a depressing effect on price.

Economic fundamentals can also induce interesting relationships between other commodities. In electricity, the spark spread between the price of power and the price of fuel (such as natural gas) used to generate it reflects the demand and supply for electricity. Although shocks to electricity demand move electricity and fuel prices in the same direction, power prices tend to move more. Moreover, Pirrong (2008a) demonstrates that the correlation between power and fuel prices is time varying; correlations are high when the demand for power is low and power prices are low, and correlations are low when the demand for power is high and power prices are high. This occurs because when demand is high, prices reflect the scarcity value of generating capacity, which does not depend on fuel prices; but when demand is low, prices are driven primarily by the marginal cost of generation, which is determined primarily by the price of fuel.

Energy prices exhibit higher volatilities than virtually any other commodity or commodity class (Duffie, Gray, Hoang 2004). For instance, the annualized volatility of front-month crude oil over the period from 1992 to 2004 period was 36 percent, while the volatility of natural gas during the same period was 52. In contrast, gold volatility was 16 percent, soybean volatility was 21 percent, and copper volatility was 24 percent. Energy price volatilities are time varying (Duffie, Gray, Hoang 2004). Moreover, there is evidence (at least from crude oil) that prices are more volatile when the market is in backwardation than when it is not and that the correlation between spot and futures prices is lower when the market is in backwardation than when it is not (Pirrong, 1995). These findings are consistent with the dynamic programming version of the theory of storage.

CLEARING

Energy derivatives are one product where clearing has made extensive penetration into the OTC market, especially in the natural gas market, and especially for gas basis swaps. Both NYMEX and ICE clear OTC energy products. The robust growth in energy clearing is attributable in large part to the severe credit problems experienced in the energy trading sector following the implosion of merchant energy firms such as Enron, Dynegy, Williams, and Mirant.

The NYMEX and ICE OTC clearing models differ. NYMEX clears deals executed either on its ClearPort trading system as well as deals executed with voice brokers or on a principal-to-principal basis. In contrast, ICE clears deals executed on its trading platform. NYMEX clears transactions through its own clearing subsidiary. Until September 2008, ICE cleared through the London Clearing House Ltd. (LCH), but it has since shifted its OTC clearing to its own clearing subsidiary. Moreover, NYMEX clears via a mechanism called Exchange of Futures for Swaps (EFS) whereby each counterparty's swap position is exchanged for an equivalent position in a look-alike futures contract; ICE deals remain OTC subsequent to clearing.

Clearing volume on NYMEX has risen from less than 30,000 contracts per day in January 2004 to over 400,000 contracts/day in June 2008.

RECENT DEVELOPMENTS

The dramatic rise in oil prices during 2007–2008 focused considerable attention on oil derivatives trading. Speculation in oil futures and swaps was widely blamed for contributing to the rapid price increases, with some commentators opining that as much as $70 per barrel of the price of oil (out of a total price of around $140/barrel) was attributable to the effects of speculation. These concerns led to myriad legislative proposals in the U.S. Congress to constrain speculation, none of which have been enacted as of the time this is being written in August 2008.

Relatedly, the breadth of participation in energy markets has increased markedly in recent years, with the entry of banks, investment banks, hedge funds, and portfolio investors (including pension funds) into the energy market. Many portfolio investors participate in the energy derivatives markets through the purchase of or, less frequently, the sale of OTC contracts with payoffs tied to commodity indices, such as the Standards & Poor's Goldman Sachs Commodity Index (S&P GSCI) or the Commodity Research Bureau Commodity Index. In turn, financial institutions that make markets in these OTC contracts lay off some of the risks they incur through the futures markets. Energy represents a large fraction of these indices. For instance, in August 2008, energy represented 76.88 percent of the S&P GSCI, with oil alone accounting for approximately 56 percent of the total index value. The predominance of long-only index investors has been widely cited as a source of increased demand for oil that has contributed to higher energy prices. This belief, in turn, has spurred legislative proposals (none enacted as of this writing) to restrict commodity index investments by financial institutions.

These antispeculation and anti-index arguments and legislative efforts echo earlier criticisms of derivatives trading in grain and share the same conceptual deficiencies. Most notably, most speculators never take delivery of physical energy and offset their derivatives positions prior to delivery; indeed, speculators using OTC contracts and index investors do not even have the contractual right to take delivery. Hence, these participants do not contribute any demand for physical energy. Moreover, there is no evidence of physical market distortions (such as the accumulation of large inventories in the hands of speculators) that would accompany price distortions. It should also be noted that the Interagency Task Force

(2008) of the U.S. government attributed the large rise in oil prices in 2007–2008 to economic fundamentals rather than speculation, as did the International Energy Agency (2008).

REFERENCES

Bessembinder, H., and M. Lemmon. 2002. "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets." *Journal of Finance* 57: 1347–1382.

Duffie, D., S. Gray, and P. Hoang. 2004. "Volatility in Energy Prices," in V. Kaminski, ed., *Managing Energy Price Risk*. London: Risk Books.

Eydeland, A., and K. Wolniac. 2004. "Power Forward Curves," in V. Kaminski, ed., *Managing Energy Price Risk*. London: Risk Books.

Hotelling, H. 1931. "The Economics of Exhaustible Resources." *Journal of Political Economy* 39: 137–175.

Interagency Task Force on Commodity Markets. 2008. "Interim Report on Crude Oil."

International Energy Agency. 2008. "Oil Market Report." Available at www.worldoilmarket report.org.

Litzenberger, R., and N. Rabinowitz. 1995. "Backwardation in Oil Futures Markets: Theory and Empirical Evidence." *Journal of Finance* 50: 1517–1545.

Pindyck, R. 1994. "Inventories and the Short Run Dynamics of Commodity Prices." *Rand Journal of Economics* 25: 141–159.

Pirrong, C. 1995. "Metallgesellschaft: A Prudent Hedger Ruined, or a Wildcatter on NYMEX?" *Journal of Futures Markets* 17: 543–578.

Pirrong, C. 2008a. "The Price of Power: The Pricing of Power Derivatives." *Journal of Banking and Finance.*Amsterdam: Elsevier.

Pirrong, C. 2008b. "Stochastic Fundamental Volatility, Speculation, and Commodity Storage." Working paper, Bauer College of Business, University of Houston.

Weiner, R. 1992. "Origins of Futures Trading: The Oil Exchanges in the 19th Century," *Cahiers du GREEN* 92–09, Université, Laval.

ABOUT THE AUTHOR

Craig Pirrong is professor of finance, and energy markets director for the Global Energy Management Institute at the Bauer College of Business of the University of Houston. He was previously Watson Family Professor of Commodity and Financial Risk Management at Oklahoma State University, and a faculty member at the University of Michigan, the University of Chicago, and Washington University. Professor Pirrong's research focuses on the organization of financial exchanges, derivatives clearing, competition between exchanges, commodity markets, the relation between market fundamentals and commodity price dynamics, and the implications of this relation for the pricing of commodity derivatives. He has published 30 articles in professional publications, is the author of three books, and has consulted widely, primarily on commodity and market manipulation–related issues. He holds a BA in economics, an MBA, and a PhD in business economics from the University of Chicago.

CHAPTER 10

Interest Rate Derivatives

IAN LANG, CFA
RBS Greenwich Capital

Around the time the Bretton Woods exchange rate system was breaking down in global markets, interest rates in many local markets became much less predictable, leaving businesses and investors exposed to far more interest rate risk than they were familiar with. The financial community began looking for ways to hedge and speculate this risk.

To address this need to hedge interest rate risk, the futures exchanges in Chicago and London (and other financial centers) began offering interest rate contracts in addition to their historically traded commodity futures. In 1981, IBM and the World Bank entered into the first interest rate swap, soon followed by a foreign exchange swap. As the Black-Scholes model and other options pricing models became generally accepted, options on interest rates became widely traded in deep, liquid markets.

Decades later, the interest rate derivatives markets, both exchange traded and over the counter (OTC), have grown into a huge global marketplace (notional amounts in billions of U.S. dollars equivalent). (See Exhibit 10.1.)

These so-called plain vanilla interest rate derivatives totaled almost U.S. $355 trillion at the end of 2006, and there are many other interest rate derivatives that the Bank for International Settlements (BIS) does not report statistics on. At first glance, the BIS total appears to be roughly 10 times global gross domestic product. However, some OTC derivatives are double counted (e.g., each side of the same risk may be on the books of two banks; in practice this is difficult to net). A significant percentage of the total (exchange traded plus OTC) is a hedge against other interest rate risk but does not get netted in the statistics just cited. Most important, the numbers shown reflect notional amounts (i.e., the amounts used to as a benchmark for determining the dollar value of interest rate payments). The market value of these instruments is much smaller.

Whole books have been written on each type of interest rate derivative, but the next section provides a sampling of the common types of interest rate derivatives you are likely to see on a broker-dealer fixed income trading floor.

EXCHANGE-TRADED (LISTED) DERIVATIVES

Exchange-traded derivatives offer market participants a number of useful features. They are traded in standardized denominations and risk characteristics, which

Exhibit 10.1 Worldwide Interest Rate Derivatives Trading Volume

(billions of US$)	December 1996	December 2006	June 2007
Exchange-traded futures	6,180	25,683	31,677
Interest rate	5,931	24,476	30,148
Exchange-traded options	3,704	44,759	66,245
Interest rate	3,278	38,117	56,454
Over the counter	22,290	414,290	516,407
Interest rate swaps, forwards, and options	19,171	291,115	346,937

Source: BIS Quarterly Review (August 1997, December 2007).

enhances liquidity and price discovery by encouraging a large number of participants. The clearinghouse acts as a counterparty to all transactions and collects risk-based margin from clearing members as well as mandating minimum margin requirements for customers of clearing members. In sum, this almost eliminates counterparty risk, the risk that one party to the transaction might default. (A contract holder is exposed to risk of the clearinghouse failing, however improbable this might be.)

Exchanges allow trading in futures contracts and options thereon. Interest rate futures are among the most liquid instruments traded, with a very small bid-ask spread (one tick in many cases). Bid-ask spreads on options are generally wider, especially for away-from-the-money strikes.

Eurodollar futures are forward-settling agreements to deposit U.S. $1 million into a London Interbank Offer Rate (LIBOR) account for 90 days starting at the contract expiry. They are the exchange-traded equivalent of a forward rate agreement. LIBOR is the rate paid by the most creditworthy banks in the global economy; less creditworthy names will pay LIBOR plus a spread. These are the most liquid futures traded, and their rates often are used to price the first two years of the swap curve. Liquid markets exist for contracts out of at least 4 and less active markets extend up to 10 years. Portfolio managers use Eurodollar futures to hedge short-term interest rates and to profit from anticipated changes in the macro economy, as short-term rates are generally more responsive to oncoming monetary easing or tightening. When there are major market dislocations, traders will instinctively buy Eurodollar futures, anticipating both a flight to quality and the possibility that central banks may ease monetary policy. Eurodollars are also used by broker-dealers to hedge interest rate risk in their swap books.

In addition to Eurodollar futures (which are for U.S.-dollar-denominated deposits), there are very liquid futures markets in Euribor (euro denominated), Short Sterling (British pound denominated) and TIEE MexDer (28-day interbank rates in Mexican pesos). Less active markets include EuroFrancs (Swiss francs), Tibor (Japanese yen denominated) futures, and bankers' acceptance futures (Canadian dollar denominated).

The other short-term interest rate future in the United States is a future on Fed Funds—or, more precisely, on the monthly average of effective Fed Funds. This contract is still very active in nearby months, especially when the Fed is expected to change monetary policy in the near future. To hedge or speculate more than a few months forward, market participants often use the Overnight Interest Swap (OIS) market (which is not exchange traded). OIS term rates are based on the effective Fed Funds rate and are actively quoted to a year or more in the future. The OIS curve is also watched by many portfolio managers to gauge what rate changes the market has already priced in. While the United States has a Fed Funds future contract, markets in most other countries rely exclusively on their local-currency-denominated OIS market for hedging central bank policy.

The Eurodollar, Euribor, Euroyen, and other short-term interest rate futures trade risk on 90-day interest rates. There are also futures traded on longer maturity interest rates. In the United States, futures are traded on 2-year, 5-year, and 10-year U.S. Treasury notes and on 30-year U.S. Treasury bonds. In Europe, futures are traded on German government debt, such as Schatz (2 year), Bobl (5 year), and Bunds (10 year). The United Kingdom trades futures on Gilts (10-year debt). Canada and Japan both have less liquid markets in 10-year bond futures.

Futures on government debt are forward-settling trades that call for delivery of one or more bonds from a delivery basket—usually all issues of the stated government whose maturity on the future delivery date will be within range of the future's stated maturity. For example, notes with a maturity from 6.5 to 10 years can be delivered to satisfy the contract obligations of the Chicago Board of Trade (CBOT) future on U.S. 10-year notes. The exchange specifies conversion factors that determine how much nominal quantity of a given note must be delivered against a futures contract. In theory, these conversion factors are supposed to make the cost of delivering any note in the deliverable basket roughly equal to any of the others. In practice, one bond is usually cheapest to deliver (CTD), and the future will trade as a forward settling trade on that one bond (plus the basis). In the cash markets, this CTD bond often will trade slightly rich to other similar maturity bonds and will trade tighter in the repo markets, as traders who are short the futures contract will want to deliver this bond against their futures position.

The futures price usually will trade slightly cheaper than a simple forward trade on the CTD (the difference is called the basis). The futures contract is a portfolio that is long the CTD bond but short several options. The seller of the futures has the option to deliver any bond in the basket (and another bond may become cheapest to deliver, depending on changes in the yield curve and repo markets). The futures seller also has an option of when to deliver, generally anytime during the contract month. This usually means that bonds with positive carry will be delivered at the end of the month, and those with negative carry will be delivered on the first business day. In a flat yield curve environment (where carry is essentially zero), the bond might be delivered at any time during the month. All these embedded options are worth more when there is greater volatility in yields. Recently, the CTD bond for most futures has been the shortest-maturity bond in the basket; however, if rates were to rise significantly, futures sellers would have the option of delivering longer-maturity (higher-duration) bonds against the contract. Hence the embedded options give a futures contract negative convexity and are short gamma trades when compared against a simple cash bond position. Please see Chapter 25 for more details.

There are very liquid markets in exchange-traded options on Eurodollar and Euribor futures as well as on many of the longer maturity futures. Please see Chapter 29 for discussion of how these options are priced and traded.

OVER-THE-COUNTER DERIVATIVES

Over-the-counter derivatives are so named because they do not trade on an exchange—they are a contract between two private entities. This means each party is potentially exposed to credit risk that the other party may not be able or willing to make good on the terms of the contract. However, they have the advantage of being customized to hedge the exact risk(s) and nominal size that the parties want to include. Exchange-traded derivatives trade only the most liquid risks and in amounts that may be too big or too small for the needs of a market participant.

OTC Options

Perhaps the most obvious, but certainly not the most liquid, OTC interest rate derivative is an option (either put or call) on a cash bond. These options usually are priced against a specific bond (not a basket like the exchange-traded options on futures). Most often, the bond is a government debt instrument, but OTC options also are traded against agency debt, mortgage-backed securities and corporate debt. OTC options are less liquid than their exchange-traded counterparts, and involve credit risk. As a result, they generally trade at higher implied volatility than exchange-traded options and have a wider bid-ask spread. Since the option is struck against a specific bond, the financing rate (and hence forward price) is subject to the availability of the underlying bond in the repo markets. No two OTC options are exactly the same, so two options with similar characteristics may be priced differently. There is no actively traded market for an existing OTC option, which in practice means it is difficult to mark these options to market on a daily basis. Because of the individuality of each option, a market maker often cannot precisely offset the risk of one option with others, which makes the cost of hedging the risk in the trade higher.

Rate Locks

Another common OTC derivative is a rate lock. When an issuer (usually a corporation) is considering selling bonds to finance its business, the treasurer would like to lock in an interest rate so that the company can determine whether the projected return on the venture will exceed the costs. When a company decides to sell debt, it can take several months to hire investment banks, print a prospectus, determine investor interest, register the securities with regulators, and then actually issue debt. Even in the case when a company has a shelf registration, there can be a substantial lag between the time a corporation finds an attractive interest rate to issue debt and when that debt is actually issued.

To lock in the attractive interest rate, a corporation could short a bond or swap of similar maturity (a hedge). If rates rise before debt is issued, the profit from being short the hedge should offset the added cost of paying a higher rate on the issued debt. If rates fall, the loss on the hedge will offset the lower rate paid. Hence selling a bond short can effectively lock in an interest rate.

In practice, shorting a bond introduces a number of problems. It may be difficult to borrow a sufficient amount of a bond to sell short (either one is not available or the borrowing costs are too high). Selling the bond means locking in a repo rate from now until an unknown future date or else continuously paying overnight repo rates, which might move adversely. Paying fixed on a swap would similarly subject the company to uncertain cash flows because the floating revenue from the swap is unknown. The company may not have a credit rating or regulatory authority to trade swaps. Next, the act of selling the bond will appear on the company's balance sheet, which may complicate other financing activities. The short is not technically part of the company's business; it is there only to lock in an attractive debt interest rate. Its presence on the balance sheet can distract creditors and investors from the company's actual business activity. Finally, a debt offering usually is made for a specified maturity: 10 years, for example. If this is hedged by selling a 10-year bond, and the issuance takes 6 months, then a 10-year debt issuance is being hedged with a 9.5-year hedge, which introduces unwanted curve risk. The profit or loss on the hedge will no longer perfectly offset a change in 10-year rates.

To address these risks, bond dealers will sell the company an interest rate lock. The interest rate lock will pay the company (or cost the company) on the contracted date an amount that offsets the cost or benefit of the company paying a different rate on its debt than what it had planned to pay. As an OTC derivative, the lock can be structured to have the exact maturity and size the corporation requires, and can be unwound early if needed. The costs of balance sheet usage, short-selling repo risks, curve risk, and interest rate risk are taken away from the corporation and taken on by the investment bank.

Swaps and Swaptions

Probably the most common OTC derivatives are swaps and swaptions (options on swaps). Swaps are an agreement between two parties to exchange (swap) cash flows—most commonly cash flows based on a fixed semiannual interest rate against cash flows based on quarterly floating rates. The rates, both the fixed rate and the series of floating rates, are based most commonly on the Eurodollar/LIBOR curve. A swap done at market rates does not require any exchange of principal, as the present value of semiannual fixed rate payments is equal to the present value of the quarterly floating rate payments. Please see Chapter 28 for more on how to value interest rate swaps.

Swaps can be thought of as a bond position with built-in financing. The fixed rate receiver earns the fixed rate (similar earning the coupon on a bond) while paying the quarterly floating rate (similar to financing the bond in the repo market). Interest rate swaps are used to hedge interest rate exposure in all types of bonds and to trade asset swaps—the spread between the swap curve and another curve. For example, when credit spreads are expected to widen, traders may buy Treasury bonds and pay fixed on a matched maturity swap.

Not every portfolio can trade swaps. Some portfolios are prohibited by law or by investment policy. Others are restricted for credit reasons: Since there is no clearinghouse involved, there is potential credit risk in each trade. As a result, swaps usually are traded only between parties with the highest credit ratings. Swap and swaption trades are governed by a standard International Swap Derivatives Association (ISDA) agreement, which needs to be signed by both parties before swap trading begins.

Swaptions are options to pay (swaption payers) or receive (swaption receivers) fixed rate on a swap. The common representation is to quote, for example, a 5X5 or a 3mX10 swaption. A 3mX10 option has a term of three months (the option term) and gives the holder the right (but not the obligation) to enter into a 10-year swap (the tail) that starts 3 months from now. A 5X5 swaption is an option to enter into a 5-year swap starting 5 years from now. Many dealers also make a market in straddles, which are a receiver and a payer swaption priced as a single trade.

Caps, Floors, and Collars

There is a liquid market for caps and floors, which can be thought of as a strip of short-term swaptions that effectively create a cap or floor on a short-term floating rate. Each of the options in the strip is called a caplet or floorlet.

For example, a corporation issues floating rate debt that pays three-month LIBOR each quarter and uses this debt to fund a business that earns 8 percent. The corporation wants to protect itself from the risk that three-month LIBOR might be higher than 8 percent. A dealer will sell the corporation a cap struck at 8 percent that will pay the corporation the difference between 8 percent and three-month LIBOR, if the three-month LIBOR is above 8 percent (and nothing if LIBOR is below 8 percent). In effect, the cap means company will pay three-month LIBOR but no more than 8 percent.

A floor is the mirror image of a cap; it pays the owner if the floating rate falls below the strike, effectively insuring that payments received will not drop below a minimum "floor" level. A collar is the combination of long a cap and short a floor, in effect keeping the floating rate inside a range. The collar often is structured with strikes such that the cap and floor have the same price, thus giving the collar a zero cost.

Dealers will quote prices for swaptions in basis points (hundredths of 1 percent) so that customers can calculate a price for the notional amount they require. From these basis point prices, one can also calculate an implied volatility that has been assumed for the swaption. A matrix of these implied volatilities, with different terms on one axis and different tails on the other, forms a volatility surface. Because swaptions are very liquid and therefore have lots of price discovery, the swaption volume surface is watched closely by traders of many other fixed income products, such as callable agency bonds, corporate bonds, and mortgage bonds.

The mortgage market in particular watches 3mX10 and 5X5 swaption volatilities carefully for indicators of what areas of the market might have better value. While traders will use a much larger selection of swaptions actually to hedge a mortgage book, they often watch 3mX10 and 5X5 as intraday indicators of what hedge changes might be needed. (The optimal hedge involves very complex calculations.)

Short-term swaptions, such as 3mX10, are indicators of how the market is pricing delta and gamma risk—the risk of a price change resulting from changes in the level of interest rates. The short term (3 months) makes these options sensitive to the market's perceptions of whether interest rate levels may change in the near future, while the long tail (10 years) is a very rough approximation of how long a typical 30-year mortgage will remain outstanding before being paid off (e.g., the homeowner moves to a new house or decides to refinance).

Longer term-swaptions, such as 5X5, are good indicators of how the market is pricing volatility (vega risk). Because of the long term (5 years), these swaptions are less sensitive to change in the level of rates but have much higher vega, which measures the swaption's price sensitivity to a change in implied volatility. Owning a mortgage bond is very similar to owning a fixed-rate bond (so traders are interested in the level of rates, just as with any other bond) and being short a strip of calls on that bond. The price of those theoretical calls one is short will increase when volatility increases—which, all else being equal, makes the price of the mortgage decline.

Holders of large mortgage portfolios, including government-sponsored entities (GSEs) such as Fannie Mae and Freddie Mac, are implicitly short a huge portfolio of calls. To manage the risk in their mortgage portfolio, mortgage portfolio owners need to manage the risk of these calls. They do this partly by buying similar swaption receivers (which have a similar profit-and-loss profile as the embedded calls). However, the mortgage market is the biggest sector of the fixed income market, and mortgage risks tend to be concentrated at known strikes (because homeowners have become more sophisticated about refinancing when rates drop). In practice, if mortgage owners were to hedge all their embedded option risk, they would give up all the extra spread (the mortgage basis) that mortgages yield above Treasuries. Mortgage risks are selectively hedged.

Mortgage Derivatives

The second common type of OTC interest rate derivative is mortgage derivatives. Until recently, a secondary market for securitized mortgages existed only in the United States, while swaps and swaptions markets existed in most developed countries. In the past 5 to 10 years, a secondary market to trade mortgage-backed securities has developed in Europe, but as of this writing the European mortgage market is smaller than that in the United States.

When mortgages are securitized, they are grouped together into pools, which are the most basic mortgage security. As described in the discussion of swaptions, mortgage securities have embedded calls that make it difficult to determine precisely when cash flows can be expected. These embedded calls are the result of the homeowners' option to prepay the mortgage at any time—to refinance, to move, or simply to reduce debt. The embedded calls give the mortgage-backed security negative convexity, but the premium received from "selling" these calls give the mortgages an added yield relative to noncallable securities.

Mortgage TBAs (to be announced) are the simplest form of mortgage derivative. They are very similar to listed futures contracts, except the underlying security is obviously a mortgage, and TBAs trade OTC; there is no clearinghouse backing the trade. In a mortgage TBA trade, the seller agrees to deliver some quantity of mortgage pools to the buyer at a specified future date. Market conventions determine the exact date as well as the characteristics of what pool(s) can be delivered. Like the futures contract, there is a basket of possible pools that are eligible for delivery, and like the futures contract, the seller will try to deliver the cheapest acceptable pools against the TBA. Nonetheless, TBAs will approximate the profit and loss performance of this unknown (to be announced) group of mortgages. Indeed, some portfolio managers choose to get much, and sometimes all, of their mortgage exposure through TBAs. Each month, near the delivery date, they roll their current TBA into the next month's TBA. Since the TBA is a forward-settled derivative on the cheapest-to-deliver, and not necessarily the best-performing pools, it follows that there are often more attractive investment opportunities in investing in individual pools. However, buying individual pools involves substantially more back office complexity, as a pool owner must track monthly cash flows and mortgage factors (the percentage of the original mortgages that have not been prepaid). Mortgage TBAs also do not require the owner to commit balance sheet, which allows both for leverage and for using cash to pursue attractive relative value opportunities that require balance sheet.

Mortgage pools are themselves grouped into bigger pools. The aggregate cash flows are then restructured to create collateralized mortgage obligation (CMO) tranches that have different risk characteristics from the original pools. Senior tranches may have more predictable (less volatile) prepayment characteristics, or they might receive all prepayments initially (effectively making them much shorter duration assets). Junior tranches then will have more prepayment risk and/or later prepayment of principal, and they generally will have higher yields to compensate for this risk. Still other tranches may pay floating interest rates, inverse floating rates, principal only, or interest only. Investment bankers create tranches with varying characteristics according to market demand.

A typical mortgage desk also will trade securities collateralized by commercial mortgages (CMBS) and asset-backed securities, which are backed by cash flow streams from consumer credit card debt; aircraft leasing; satellite leasing; automobile loans and leasing; container, railcar, or other equipment leasing; accounts receivable/factoring; and many other sources.

FURTHER READING

Burghardt, G. 2003. *The Eurodollars Futures and Options Handbook*. New York: McGraw-Hill.
Burghardt, G. 2005. *The Treasury Bond Basis*. New York: McGraw-Hill.
Kolb, R. and J. Overdahl. 2007. *Futures, Options and Swaps*. Hoboken, NJ: John Wiley & Sons.
Hull, J. C. 2005. *Options, Futures and Other Derivatives*, 6th ed. Saddle River, NJ: Prentice Hall.
Fabozzi, F. 2005. *The Handbook of Fixed Income Securities*, 7th ed. New York: McGraw-Hill.

ABOUT THE AUTHOR

Ian Lang was a senior vice president at RBS Greenwich Capital Markets while writing his chapter and has since moved to a private buy-side firm. While at Greenwich Capital, Mr. Lang held a number of positions where he was responsible for developing and maintaining proprietary analytics for listed and over the counter securities, as well as providing customer education/documentation. Prior to Greenwich Capital, he developed mortgage CMO analytics at Bloomberg LP. Mr. Lang earned a BA in economics in 1992 from the University of Rochester and is a CFA charter holder.

CHAPTER 11

Exotic Options

ROBERT W. KOLB
Professor of Finance and Frank W. Considine Chair of Applied Ethics,
Loyola University Chicago

OVERVIEW

In recent years, financial engineers have created a variety of complex options that are collectively known as exotic options, which we can contrast with plain vanilla options. For a plain vanilla option, the value of an option at any particular moment depends only on the current price of the underlying good, the exercise price, the risk-free rate of interest, the volatility of the underlying good, the time until expiration, and the dividend rate on the underlying good. Further, there is a fixed underlying good, a fixed and stated exercise price, and a known time to expiration, and there are no special conditions on any of the option parameters. Perhaps most important, plain vanilla options have a profit and loss function at expiration that has the familiar "hockey stick" graph. That is, for a call option the option has a total loss of the premium for any stock price at expiration equal to the exercise price or less, and then with the profit increasing dollar for dollar without limit as the stock price exceeds the exercise price. Exotic options, by contrast, have a variety of very different payoff patterns and characteristics.

The price of a plain vanilla option depends only on the current price of the underlying good and not the history of that price. Therefore, the price of a plain vanilla option is independent of the price path followed by the underlying good. In contrast, many exotic options exhibit path dependence: The price of the option today depends on the previous or future price path followed by the underlying good. For example, the price of a lookback call option depends on the minimum price reached by the underlying good over some past period. Further, the price of an average price option depends on the future average price of the underlying good. Thus, to price a path-dependent option, it is not enough to know the current price of the underlying good. Instead, we must have information about the previous path that the price of the underlying good traversed.[1]

This chapter discusses nine of the most common exotic options:

1. Forward-start options
2. Compound options
3. Chooser options
4. Barrier options
5. Binary options

6. Lookback options
7. Average price options
8. Exchange options
9. Rainbow options

Because of the complexity of these options, we emphasize European options, and the chapter explains the basic payoff conditions and other terms of these options. Nonetheless, it is important to realize that the catalog of exotic options continues to grow as financial engineers construct ever more innovative options with widely varying and increasingly complex payoff patterns.

For most American exotic options, and for some European exotic options, there is no exact pricing formula. For these options, simulation or approximation methods must be used to estimate the price. This process adds considerable complexity and is beyond the scope of this chapter. For the complete valuation formulas for the options discussed in this chapter, see Kolb and Overdahl (2007), Chapter 18, or Zhang (1998).

FORWARD-START OPTIONS

One of the simplest of exotic options is a forward-start option, in which the price of the option is paid at the present but the life of the option starts at a future date. The exercise price typically is specified to be the current price at the beginning of the option's life—that is, the option contract specifies that the option will be at-the-money when the option's life begins. Forward-start call options are often used in executive compensation packages. An executive might receive a forward-start call option on the firm's shares with an exercise price to equal the firm's share price at the time the option life starts. Therefore, a forward-start option today is essentially a deferred granting of an option with a stock and exercise price equal to today's stock price and a time to expiration that equals the period from the grant date to the final expiration date.[2]

COMPOUND OPTIONS

A compound option is an option on an option; in other words, when one option is exercised, the underlying good is another option. There are four types of compound options: a call on a call, a call on a put, a put on a call, and a put on a put. For example, consider the owner of a call on a call. The owner of the compound call has until the expiration date of the compound option (the call on a call) to decide whether to exercise the compound option. If she exercises, she receives the underlying call option with its own exercise price and time until expiration. If she subsequently exercises that underlying option, she receives the underlying good.

For a European option, the owner of the compound option cannot exercise until the expiration date of the compound option. If he exercises the compound option, he will immediately receive the underlying call in the case we are considering. Therefore, when the compound option is at expiration, the choice is really very simple: Pay the exercise price of the compound option and receive the underlying option, or do nothing and allow the option to expire worthless. Thus, when the compound option reaches expiration, the trader will exercise the compound call if the price of the underlying call is worth more than the exercise price of the compound option. At the expiration date of the compound option, the underlying call (or underlying put) can be priced according to the Black-Scholes-Merton model.

Before the expiration date of the compound option, its value depends on the value of the underlying good in a two ways.

1. The value of the underlying option is largely a function of the value of the underlying good, as it is for all options.
2. The value of the compound option also depends on the price of its underlying good, which is the option underlying the compound option.

The valuation of these compound options is highly analogous to the valuation of an American option with a dividend payment between the valuation date and the expiration date.[3] The key in valuing such an option is to find the critical stock price that makes the owner of the underlying call indifferent between exercising and allowing the option to expire worthless. In the case of an option on an underlying call, the critical stock price will be the stock price that leaves the owner of the underlying call indifferent between exercising or not. Therefore, for a compound option on an underlying call, the critical price is the stock price at which the value of the underlying call equals the cost of acquiring it.

CHOOSER OPTIONS

The owner of a chooser option has the right to determine whether the chooser option will become a call or a put option by a specified choice date. After the choice date, the resulting option is a plain vanilla call or put, depending on the owner's choice. A chooser option is also known as an as-you-like-it option. Chooser options are useful for hedging a future event that might not occur. For example, assume Congress is debating a trade bill with potentially large implications for the value of the foreign currency of a trading partner. If the bill passes, the presumption is that the value of the foreign currency will increase, but the bill's defeat is expected to be negative for the value of the foreign currency. Traders could hedge this uncertainty with a chooser option on the foreign currency. If the trade bill passes, the owner can choose to let the option be a call on the foreign currency; if the bill fails, the owner can choose to let the option be a put.[4]

In considering chooser options, there are three dates to consider: the valuation date; the choice date, when the owner of the chooser must choose for the option to be a call or put; and the expiration of the option. At the choice date, the option is chosen to be a call or a put, and it can be valued by the Black-Scholes-Merton model at that time. However, for the chooser option, the problem is to evaluate the option before the choice date. There are simple and complex chooser options. For a simple chooser option, the potential put and call have a common exercise price and expiration date. Complex choosers allow the potential call and put to have different exercise prices, different expiration dates, or both different exercise prices and expiration dates.

BARRIER OPTIONS

Barrier options can be "in" options or "out" options. An "in" barrier option has no value until the price of the underlying good touches a certain barrier price. When that happens, the option becomes a plain vanilla option. Accordingly, an "out" option is initially like a plain vanilla option, except if the price of the underlying good penetrates the stated barrier, the option immediately expires worthless. Barrier options can be either calls or puts, permitting eight types of barrier options:

1. Down-and-in call
2. Up-and-in call
3. Down-and-in put
4. Up-and-in put
5. Down-and-out call
6. Up-and-out call
7. Down-and-out put
8. Up-and-out put

Barrier options may also pay a rebate, which is a booby prize. For an "out" barrier option, the rebate is paid immediately when the barrier is hit and the option passes out of existence. For an "in" barrier option, the rebate is paid if the option expires without ever hitting the barrier price. Barrier options are also known as knock-in and knock-out options, and these options exhibit path dependence. The value of a barrier option at the present depends on the previous sequence of stock prices, particularly on whether the stock price has already hit the barrier for an "in" option. The present price of a barrier option can also depend on the future price path of the underlying stock: Will the price hit the barrier between now and expiration?[5]

Barrier options may be viewed as conditional plain vanilla options. "In" barrier options become plain vanilla options if the barrier is hit. "Out" barrier options are plain vanilla options, with the condition that they may pass out of existence if the barrier is hit. These conditions make barrier options inferior to unconditional plain vanilla options, so barrier options will be cheaper than otherwise identical plain vanilla options. This cheapness gives barrier options a special usefulness in hedging applications. For example, a portfolio manager may expect the value of her portfolio to increase but wishes to protect against the possibility of a large drop in value. Accordingly, she might buy a put option with an exercise price slightly below the current price of the portfolio. This is essentially a portfolio insurance strategy. By buying a down-and-in put instead of a plain vanilla put, the portfolio manager can get similar protection but at a cheaper price.

Exhibit 11.1 shows how payoffs arise for a down-and-in option. As mentioned earlier, the stock price must be above the barrier for such an option to be interesting; if the stock price is below the barrier, the barrier has been touched and the barrier option has already become a plain vanilla option. Thus, the initial stock price will exceed the barrier, but the exercise price may be either higher or lower than the barrier. The top panel of Exhibit 11.1 shows a situation in which the

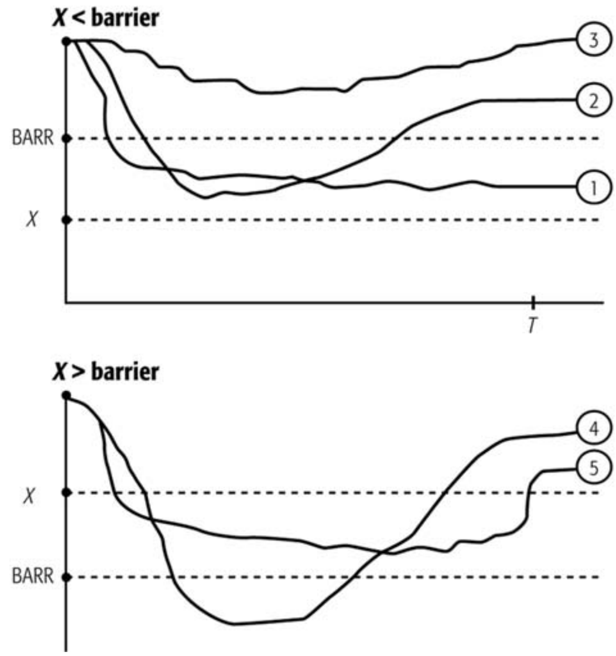


Exhibit 11.1 Alternative Stock Price Paths for Down-and-in Barrier Options

barrier, BARR, exceeds the exercise price, *X*. Three stock price paths are shown in this panel:

1. The stock price may penetrate the barrier and terminate above the exercise price but below the barrier.
2. The stock price may penetrate the barrier, and the terminal price might exceed the barrier. (We distinguish these two cases, because the probabilities associated with these two price paths are different, even though they have the same payoff.)
3. The stock price may never penetrate the barrier, and the payoff is the rebate amount, REBATE.

The bottom panel shows a similar situation, except in this case the exercise price exceeds the barrier. In price path 4, the stock price penetrates the barrier, but terminates above the exercise price. Finally, in price path 5, the barrier is never hit, so the payoff is REBATE. As Exhibit 11.1 pertains only to down-and-in options, we must also consider down-and-out options, particularly the fact that the rebate is paid immediately upon the barrier being pierced. This gives a sixth payoff possibility. The value of a down-and-in call is the present expected value of these payoffs.

BINARY OPTIONS

Binary options have payoffs that are discontinuous, paying either nothing or a considerable amount, depending on the satisfaction of some condition. For example, a cash-or-nothing call is a type of binary option that pays a fixed cash amount if the stock price terminates above the exercise price or pays nothing if the terminal stock price is below the exercise price. These binary options are also known as digital options, a name that reflects the all-or-nothing character of their payoffs.[6] Other types of binary options that we consider in this section are asset-or-nothing options and supershares.

A cash-or-nothing call pays a fixed cash amount if the terminal stock price exceeds the exercise price; otherwise the call pays nothing. Similarly, a cash-or-nothing put pays a fixed cash amount if the terminal stock price is below the exercise price. These options require no payment of an exercise price. Instead, the exercise price merely determines whether the option owner receives a payoff. Viewed from the perspective of the valuation date, the value of a cash-or-nothing call will simply be the present value of the fixed cash payoff multiplied by the probability that the terminal stock price will exceed the exercise price. The valuation of these binary options depends critically on the probability of the option finishing in the money, and this probability is given in the Black-Scholes-Merton model.

In August 2003, a furor erupted in Washington, D.C., over the disclosure that the Pentagon intended to operate an experimental market that would have allowed online traders to trade on the probability of future terrorist attacks and the occurrence of future political events in the Middle East. For a few days, it was the best-known market in America, even though it was dissolved before trading actually commenced. Formally, it was called the Policy Analysis Market (PAM), but many news accounts referred to the market as the "turmoil exchange." The market would have traded 24 hours a day, seven days a week. The Pentagon wanted to create the PAM in order to gather information that would help validate other sources of intelligence that could be used to stop terrorism and reduce political instability.

The PAM was inspired by the Iowa Electronic Markets (IEM), where investors trade contracts on election outcomes. In over 15 years of experience, research has demonstrated that IEM markets do a better job of predicting actual election results than polls do. PAM was supposed to harness the power of markets to aggregate the knowledge and information of thousands of investors. Markets do a good job of aggregating information. Because traders are using real money to back their opinions, they have an incentive to use cold, hard, honest logic and reliable information in basing their trades. This feature mitigates the so-called yes-man effect within an organization, whereby an analyst may be tempted to tell the boss what he wants to hear. News reports cited this feature as one reason why the Pentagon was contemplating a market to aggregate intelligence information.

The PAM and IEM are part of a broader set of markets called event markets. Event markets allow participants to profit from the occurrence of a specific event. These markets go by a variety of names: nonprice markets, prediction markets, decision markets, proposition markets, opinion markets, information markets, and nontraditional markets. Several facilities offer contracts on various types of events. For example, TradeSports, based in Ireland, offers over 1,300 contracts on everything from sporting events to elections, to the probability of key terrorists being "neutralized" by a particular date. During the furor about the PAM in August 2003, they even listed a contract on the resignation of the Pentagon official in charge of the market. Other event markets, such as the Hollywood Stock Exchange, offer contracts on the box-office success of new film releases.

Event contracts are more often styled as options. The contracts typically are crafted with a fixed payout if the event occurs and zero payout if the event does not occur. A payoff structure of this sort resembles a binary option. For example, a contract may be structured so that it pays $1.00 if a Republican is elected president and zero otherwise. If the contract is currently trading at 60 cents, this means that an investment of 60 cents today could yield $1.00 if a Republican is elected. If one ignores for the moment the time value of money, the price can be interpreted as the market's assessment of the odds of the event occurring. In this case, the contract price of 60 cents can be interpreted as a 60 percent chance that a Republican will be elected.

Asset-or-nothing options are similar to cash-or-nothing options, with one major difference. Instead of paying a fixed cash amount as cash-or-nothing options do, the payoff on an asset-or-nothing option is the underlying asset. If the terminal asset price exceeds the exercise price, the owner of a call receives the asset, but if the terminal asset price is below the exercise price, the call expires worthless. For a put, if the terminal asset price is less than the exercise price, the put owner receives the asset, but if the terminal asset price exceeds the exercise price, the put expires worthless. As with cash-or-nothing options, the exercise price is never paid. Instead, the value of the asset relative to the exercise price determines whether the option pays off or is worthless.

For an asset-or-nothing call, the value is simply the present value of the asset, depreciated for dividends between the present and expiration, multiplied by the probability that the terminal asset price will exceed the exercise price. Similarly, the asset-or-nothing put is worth the present value of the asset, discounted for the dividends between the present and expiration, multiplied by the probability that the terminal asset price will be below the exercise price. A portfolio of an asset-or-nothing call and put, with the same term to expiration and underlying asset, is worth the present value of the asset discounted for the dividends to be paid over the life of the option.

A supershare is a financial instrument whose value depends on an underlying portfolio of other financial assets. A supershare represents a contingent claim on a fraction of the underlying portfolio. The contingency is that the value of the underlying portfolio must lie between a lower and an upper bound on a certain future date. If the value of the underlying portfolio lies between the bounds, the supershare is worth a proportion of the portfolio. If the value of the portfolio lies outside the bounds, the supershare expires worthless.[7]

The basic idea behind supershares is the creation of a financial intermediary that holds a portfolio of securities and issues two kinds of claims against that portfolio. The first kind of claim is a supershare, which has an uncertain payoff depending on the performance of the portfolio. The second kind of claim is a purchasing power bond that pays a given rate of real interest. A supershare is essentially like a portfolio of two asset-or-nothing calls, in which the owner of a supershare purchases an asset-or-nothing call with a low exercise price and sells an asset-or-nothing call with a higher exercise price.

LOOKBACK OPTIONS

For a plain vanilla option, the payoff depends only on the terminal stock price, not the price at any other time. For a lookback option, the exercise price and the option's payoff are functions of the price of the underlying good up to the expiration of the option. For lookback calls, the exercise price is the minimum stock price experienced over the life of the option. For lookback puts, the exercise price is the maximum stock price over the same period. Thus, it is said of lookback options that they allow the option owner to buy at the low and sell at the high. Of course, this opportunity will be priced in a rational market.

Consider the decision to purchase a lookback at time t with expiration at time T. With S_t representing the stock price at time t, the payoffs on the lookback call (LBC) and put (LBP) would be:

$$\text{LBC: MAX}\{0, S_T - \text{MIN}[S_t, S_{t+1}, \ldots, S_T]\}$$
$$\text{LBP: MAX}\{0, \text{MAX}[S_t, S_{t+1}, \ldots, S_T] - S_T\}$$

In effect, a lookback call allows the purchaser to acquire the asset at its minimum price over the life of the option, while the lookback put allows the owner to sell the asset at its maximum price over the relevant interval. Of course, the option to make these transactions has considerable value over and above an analogous plain vanilla option. Notice that lookbacks should always be exercised. For a call, the terminal stock price always will exceed some price experienced on the asset during its life. For a put, the terminal stock price will always be less than some stock price during the interval. Lookback options are clearly path-dependent options, because the value ultimately depends on the minimum or maximum stock price reached over the life of the option, not merely on the terminal price when the option expires.[8]

Because lookbacks offer cheap exercise prices for calls and high payoffs for puts, lookbacks are worth considerably more than their plain vanilla counterparts. The high premiums on lookback options have hindered their popularity in actual markets. This limitation has led to the creation of partial lookback options. These partial lookbacks restrict the minimum or maximum used in computing the payoff in some way.[9]

ASIAN OR AVERAGE PRICE OPTIONS

An Asian option is an option whose payoff depends on the average of the price of the underlying good or the average of the exercise price. (These options are called Asian options because Bankers Trust was the first to offer such products, and they offered them initially in their Tokyo office.[10]) In this section, we consider one type of Asian option, an average price option. In an average price option, the average price of the underlying good essentially takes the place of the terminal price of underlying good in determining the payoff.

Asian options are extremely useful in combating price manipulations. For example, consider a corporate executive given options on the firm's shares as part of his compensation. If the option payoff were determined by the price of the firm's shares on a particular day, the executive could enrich himself by manipulating the price of his shares for that single day. However, if the payoff of the option depended on the average closing price of the shares over a six-month period, it would be much more difficult for him to profit from a manipulation. Asian options were first used in this kind of application. As a further example, commodity-linked bonds have two forms of payoffs, the payoffs from a straight bond plus an option on the average price of the linked commodity. By making the payoff depend on the average price of the commodity, such as oil, the chance of a manipulation is lessened.[11]

The average price may be computed as either a geometric average or an arithmetic average. Unfortunately, there is no closed-form solution for the price of an arithmetic average price option, even though most actual average price options are based on an average price. These options must be valued by simulation techniques. It is possible, however, to compute the value of a geometric average price option.

A fundamental concern is the frequency with which the price will be observed over the averaging period. If the price is observed at the close each day, then the geometric average price will be computed by multiplying the available n daily price observations together and then taking the nth root of the product. An average price option may exist with some of the averaging period already under way. Alternatively, the time for averaging may lie in the future. It is typical for the averaging period to last until the option expires.

EXCHANGE OPTIONS

We now consider an option to exchange one asset for another. Upon exercising, the owned asset is exchanged for the acquired asset. The valuation of an exchange option depends on the usual parameters for the individual assets: price, risk, and dividend rate. In addition, the time until expiration and the correlation of returns between the assets also affect the valuation. We will treat the owned asset as asset 1 and the asset to be acquired as asset 2. Thus, an exchange option may be regarded as a call on asset 2, with the exercise price being the future value of asset 1.

Although exchange options were first priced in 1978, these options have existed for quite some time in the form of incentive fee arrangements, margin accounts, exchange offers, and standby commitments.[12] As an example, consider an example from the merger market. A target firm is offered the opportunity to exchange shares from the target firm for shares in the acquiring firm. The shareholders in the target firm now hold an exchange option to exchange their shares for those of the acquirer. The value of this option can range from zero to the quite valuable.

RAINBOW OPTIONS

This section considers a class of exotic options known as rainbow options. The discussion here is limited to "two-color" rainbow options—options on two risky assets, where the number of risky assets is the number of colors in the rainbow. This section discusses five types of two-color rainbow options: the best of two risky assets and a fixed cash amount, the better of two risky assets, the worse of two risky assets, the maximum of two risky assets, and the minimum of two risky assets. We consider each of these in turn, starting with an option on the best of two risky assets and cash. As we will see, the other rainbow options can be understood largely in terms of this first option.

As an example of a two-color rainbow option, consider a zero-coupon bond that pays a stated rate of interest but allows the owner of the bond to choose the currency in which the interest is paid. The value of the bond upon maturity will differ depending on the exchange rate. The right to choose the currency of repayment gives the holder of the bond a call on the maximum of two assets—the repayment in one currency or another. By contrast, consider the same type of bond but assume that the firm may choose the currency of repayment.[13]

The owner of a call option on the best of two risky assets and cash has a choice among three payoffs at expiration: risky asset 1, risky asset 2, or a fixed cash amount. There is no exercise price. Prior to exercise, the value of the option will equal the sum of the present value of these expected payoffs. Thus, the evaluation of the option turns on assessing how high the stock prices are likely to go and which asset is likely to have the highest price at expiration. The performance of the two assets will depend in part on the degree to which they are correlated.

Another type of rainbow option is a call on the maximum of two risky assets. This is similar to the option on the maximum of two risky assets and cash that we just considered. However, for this option, there is no potential cash payoff. Further, these options have an exercise price. To exercise the call, the owner pays the exercise price and selects the better of the two risky assets.

A call on the better of two risky assets is a special case of a call on two risky assets and cash. To form the special case, just specify that the exercise price is zero. The valuation of a put on the maximum of two risky assets can be derived as a function of the value of the options we have just been studying. To exercise a put on the maximum of two risky assets, the owner surrenders the more valuable of the two risky assets and receives the exercise price. Thus, the payoff from this portfolio will be exactly like the payoff from the put on the maximum of two risky assets. Therefore, the put and this portfolio must have the same value. A call on the minimum of two risky assets pays the value of the inferior risky asset upon payment of the exercise price. Upon expiration, the portfolio owner exercises the call on the more valuable asset and uses this asset to satisfy the call on the maximum that was sold to form the portfolio. These transactions have a net zero cash flow, because the portfolio owner receives and pays a common exercise price. The portfolio owner still holds the call on the inferior asset. If the inferior asset is worth the exercise price or less, the option expires worthless. If the inferior asset is worth more than the exercise price, the portfolio owner exercises for a profit equal to the difference. A call on the worse of two risky assets has a payoff equal to the inferior asset, but without the payment of any exercise price. A put on the minimum of two risky assets pays the exercise price and requires the delivery of the inferior asset.

CONCLUSION

In this chapter, we have explored a variety of exotic options. The analysis has focused on European options, for which closed-form solutions exist. As we have seen, many of these exotic options can be understood in terms of the familiar plain vanilla options priced in the Black-Scholes-Merton model.

This chapter considered nine classes of exotic options: forward-start options, compound options, chooser options, barrier options, binary options, lookback options, average price options, exchange options, and rainbow options. For each of these options, the payoffs are more complicated than those of plain vanilla options. We have seen that these specialized payoffs can be used to manage risks or to shape a speculative position more exactly. Many of these options exhibit path dependence, with the price of the option at a given time depending on the price history or the price future of the underlying asset.

Even though we have considered nine types of exotic options, this chapter has barely scratched the surface of exotic options. There are many types of exotics that are of interest that we have not even considered, such as spread option, Bermudan options, range options, basked options, mountain range options, Himalayan options, outperformance options, quantos, composite options, clique-ratchet options, and Parisian options.

ENDNOTES

1. For a good introduction to the idea of path dependence in option pricing, see W. Hunter and D. Stowe, "Path-Dependent Options: Valuation and Applications," Economic Review, Federal Reserve Bank of Atlanta, July/August 1992, pp. 30–43.
2. For more on the pricing and applications of forward-start options, see: Mark Rubinstein, "Pay Now, Choose Later," Risk, February 1991; Rubinstein and Reiner, "Exotic Options"; and Peter G. Zhang, *Exotic Options: A Guide to the Second-Generation Options*, River Edge, NJ: World Scientific Press, 1997.
3. For the original paper on pricing compound options, see R. Geske, "The Valuation of Compound Options," *Journal of Financial Economics*, 7, March 1979, pp. 63–81. See also Mark Rubinstein, "Double Trouble," *Risk*, December 1991–January 1992; Rubinstein and Reiner, "Exotic Options"; Alan Tucker, "Exotic Options," Working paper, Pace University, New York, 1995; and Zhang, *Exotic Options*.
4. See Mark Rubinstein, "Options for the Undecided," *Risk*, April 1991, and Rubinstein and Reiner, "Exotic Options."
5. For more detailed discussion of the pricing of barrier options, see: Mark Rubinstein, "Breaking Down the Barriers," *Risk*, September 1991; Rubinstein and Reiner, "Exotic Options"; Tucker, "Exotic Options"; and Zhang, *Exotic Options*. See also Emanuel Derman and Iraz Kani, "The Ins and Outs of Barrier Options," *Derivatives Quarterly*, 3:2, Winter 1996, pp. 55–67.
6. For a discussion of the pricing of binary options, see: Mark Rubinstein, "Unscrambling the Binary Code," *Risk*, October 1991; Rubinstein and Reiner, "Exotic Options"; Tucker, "Exotic Options"; Zhang, *Exotic Options*.
7. Supershares were created by Nils Hakansson, "The Purchasing Power Fund: A New Kind of Financial Intermediary," *Financial Analysts Journal*, 32, November/December 1976, pp. 49–59.
8. The first paper on lookback options appeared in 1979, long before such options actually existed. See Barry Goldman, Howard Sosin, and Mary Ann Gatto, "Path Dependent Options: Buy at the Low, Sell at the High," *Journal of Finance*, 34, December 1979, pp. 1111–27. The results of Goldman, Sosin, and Gatto were generalized to embrace a dividend paying underlying asset by Mark Garman, in his paper, "Recollection in Tranquility," *Risk*, March 1989, pp. 16–18. For more on the pricing of lookbacks, see also Rubinstein and Reiner, "Exotic Options"; Tucker.
9. For a discussion of partial lookback options, see Zhang, *Exotic Options*.
10. Zhang, *Exotic Options*.
11. For a discussion of Asian options, see: A. Kemna and A. Vorst, "A Pricing Method for Options Based on Average Asset Values," *Journal of Banking and Finance*, 14, March 1990, pp. 113–29; Rubinstein and Reiner, "Exotic Options"; S. Turnbull and L. Wakeman, "A Quick Algorithm for Pricing European Average Options," *Journal of Financial and Quantitative Analysis*, 26, September 1991, pp. 377–89; Tucker, "Exotic Options"; and

Zhang, *Exotic Options*. Kemna and Vorst give examples of several commodity-linked bonds.

12. The first paper on exchange options was by William Margrabe, "The Value of an Option to Exchange One Asset for Another," *Journal of Finance*, March 1978. Margrabe distinguished the four applications just mentioned. For additional insights on pricing exchange options, see: Mark Rubinstein, "One for Another," *Risk*, July 1991; Rubinstein and Reiner, "Exotic Options"; Tucker, "Exotic Options"; and Zhang, *Exotic Options*.
13. The original paper on rainbow options, by Rene Stulz, was "Options on the Minimum or the Maximum of Two Risky Assets," *Journal of Financial Economics*, 10, July 1982, pp. 161–85. Thus, Stulz was pricing two-color rainbow options. Stulz's work was extended to multicolored rainbow options by Herb Johnson, "Options on the Maximum or the Minimum of Several Assets," *Journal of Financial and Quantitative Analysis*, 22, September 1987, pp. 277–83. The name "rainbow option" was originated by Mark Rubinstein, "Somewhere Over the Rainbow," *Risk*, November 1991. For additional discussion of rainbow options, see: Mark Rubinstein, "Return to Oz," *Risk*, November 1994; Rubinstein and Reiner, "Exotic Options"; Tucker, "Exotic Options"; and Zhang, *Exotic Options*.

REFERENCES

Geske, Robert, "The Valuation of Compound Options," *Journal of Financial Economics*, 7, March 1979, pp. 63–81.

Goldman, Barry, Howard Sosin, and Mary Ann Gatto, "Path Dependent Options: Buy at the Low, Sell at the High," *Journal of Finance*, 34, December 1979, pp. 1111–27.

Hakansson, Nils, "The Purchasing Power Fund: A New Kind of Financial Intermediary," *Financial Analysts Journal*, 32, November/December 1976, pp. 49–59.

Hunter, W. and D. Stowe, " Path-Dependent Options: Valuation and Applications," *Economic Review*, Federal Reserve Bank of Atlanta, July/August 1992, pp. 30–43.

Johnson, Herb, "Options on the Maximum or the Minimum of Several Assets," *Journal of Financial and Quantitative Analysis*, 22, September 1987, pp. 277–83.

Kolb, Robert W. and James A. Overdahl, *Futures, Options, and Swaps 5e*, Blackwell Publishing Co., Inc. 2007.

Margrabe, William "The Value of an Option to Exchange One Asset for Another," *Journal of Finance*, March 1978.

Rubinstein, Mark and Eric Reiner, "Exotic Options," Working paper, University of California at Berkeley, 1995.

Stulz, Rene was "Options on the Minimum or the Maximum of Two Risky Assets," *Journal of Financial Economics*, 10, July 1982, pp. 161–85.

Turnbull, S. and L. Wakeman, "A Quick Algorithm for Pricing European Average Options," *Journal of Financial and Quantitative Analysis*, 26, September 1991, pp. 377–89.

Vorst, Kemna and A. Vorst, "A Pricing Method for Options Based on Average Asset Values," *Journal of Banking and Finance*, 14, March 1990, pp. 113–29.

Zhang, Peter G., *Exotic Options: A Guide to the Second-Generation Options 2e*, River Edge, NJ: World Scientific Press, 1998.

ABOUT THE AUTHOR

Robert W. Kolb is professor of finance at Loyola University Chicago, where he also holds the Frank W. Considine Chair of Applied Ethics. Dr. Kolb received two PhDs from the University of North Carolina at Chapel Hill in philosophy and finance (philosophy 1974, finance 1978). Previously he held full-time academic positions at the University of Florida, Emory University, the University of Miami, and the University of Colorado. Over his career, he has published more than 50 academic research articles and more than 20 books, most focusing on financial derivatives and their applications to risk management. In 1990, he founded Kolb Publishing Company to publish finance and economics university texts. He sold the firm to Blackwell Publishers of Oxford, England, in 1995.

CHAPTER 12

Event Derivatives

JUSTIN WOLFERS
Associate Professor of Business and Public Policy, The Wharton School, University of Pennsylvania, and CEPR, CESifo, IZA, and NBER

ERIC ZITZEWITZ
Associate Professor of Economics, Dartmouth College

In July 2003, press reports began to surface of a project within the Defense Advanced Research Projects Agency (DARPA), a research think tank within the Department of Defense, to establish a Policy Analysis Market that would allow trading in various forms of geopolitical risk. Proposed contracts were based on indices of economic health, civil stability, military disposition, conflict indicators, and potentially even specific events. For example, contracts might have been based on questions like "How fast will the non-oil output of Egypt grow next year?" or "Will the U.S. military withdraw from country A in two years or less?" Moreover, the exchange would have offered combinations of contracts, perhaps combining an economic event and a political event. The concept was to discover whether trading in such contracts could help to predict future events, and how connections between events were perceived. However, a political uproar followed. Critics savaged DARPA for proposing "terrorism futures." Rather than spend political capital defending a tiny program, the proposal was dropped.[1]

Ironically, the aftermath of the DARPA controversy provided a vivid illustration of the power of markets to provide information about probabilities of future events. An offshore betting exchange, InTrade.com, listed a new security that would pay $100 if the head of DARPA, Admiral John Poindexter, was ousted by the end of August 2003. Early trading suggested a likelihood of resignation by the end of August of 40 percent, and price fluctuations reflected ongoing news developments. Around lunchtime on July 31, reports started citing credible Pentagon insiders who claimed knowledge of an impending resignation. Within minutes of this news first surfacing (and hours before it became widely known), the price spiked to around 80. These reports left the date of Poindexter's proposed departure uncertain, which explains the remaining risk. As August dragged on, the price slowly fell back toward 50. On August 12, Poindexter issued a letter of resignation suggesting that he would resign on August 29. On August 12, the market rose sharply, closing at a price of 96.

This anecdote describes a new—and emerging—form of financial market, often known as a prediction market, but also going by the name information market or event futures. Analytically, these are markets where participants trade in contracts whose payoff depends on unknown future events. Much of the enthusiasm for prediction markets derives from the efficient markets hypothesis. In a truly efficient prediction market, the market price will be the best predictor of the event and no combination of available polls or other information can be used to improve on the market-generated forecasts. This statement does not require that all individuals in a market be rational, as long as the marginal trade in the market is motivated by rational traders. Of course, it is unlikely that prediction markets are literally efficient, but a number of successes in these markets, both with regard to public events like presidential elections and within firms, have generated substantial interest.

Although markets designed specifically for information aggregation and revelation are our focus in this chapter, the line between these kinds of prediction markets and the full range of contingent commodities—from stock in your employer's company to betting on the Super Bowl—can become blurry. However, in this chapter we generally lean away from discussing markets where the primary focus is holding or trading risk that may be intrinsically enjoyable, as in sports betting and other gambling markets. We also lean away from focusing on markets that are substantial enough in size to allow a significant extent of risk sharing and pooling by matching risky assets with risk-acceptant investors, such as the major financial markets.[2] However, most contingent commodity markets involve some mix of risk sharing, fun, and information transmission, so these distinctions are not impermeable.

We begin by describing the types of contracts that might be traded in prediction markets, before proceeding to survey several applications. We then draw together a rough and fairly optimistic description of what we have learned from early experiments, raise some market design issues, and conclude with some evidence on the limitations of prediction markets.

TYPES OF PREDICTION MARKETS

In a prediction market, payoffs are tied to unknown future events. The design of how the payoff is linked to the future event can elicit the market's expectations of a range of different parameters. We will speak as if the market is itself a representative "person" with a set of expectations. However, the reader should be warned that there are important but subtle differences between, say, the market's median expectation and the median expectation of market participants.

Exhibit 12.1 summarizes the three main types of contracts. First, in a "winner-take-all" contract, the contract costs some amount $$p$ and pays off, say, $1 if and only if a specific event occurs, such as a particular candidate winning an election. The price on a winner-take-all market represents the market's expectation of the probability that an event will occur (assuming risk neutrality).[3]

Second, in an "index" contract, the amount that the contract pays varies in a continuous way based on a number that rises or falls, such the percentage of the vote received by a candidate. The price for such a contract represents the mean value that the market assigns to the outcome.

Finally, in "spread" betting, traders differentiate themselves by bidding on the cutoff that determines whether an event occurs, such as whether a candidate

Exhibit 12.1 Contract Types: Estimating Uncertain Quantities or Probabilities

Contract	Example	Details	Reveals Market Expectation of . . .
Winner take all	Event y: Al Gore wins the popular vote.	Contract costs \$$p$. Pays \$1 if and only if event y occurs. Bid according to value of \$$p$.	Probability that event y occurs, $p(y)$.
Index	Contract pays \$1 for every percentage point of the popular vote won by Al Gore.	Contract pays \$$y$.	Mean value of outcome y: $E[y]$
Spread	Contract pays even money if Gore wins more than y^*% of the popular vote.	Contract costs \$1. Pays \$2 if $y > y^*$. Pays \$0 otherwise. Bid according to the value of y^*.	Median value of y.

receives more than a certain percentage of the popular vote. Another example of spread betting is point-spread betting in football, where the bet is either that one team will win by at least a certain number of points, or not. In spread betting, the price of the bet is fixed, but the size of the spread can adjust. When spread betting is combined with an even-money bet (i.e., winners double their money while losers receive zero), the outcome can yield the market's expectation of the median outcome, because this is only a fair bet if a payoff is as likely to occur as not.

The basic forms of these relevant contracts will reveal the market's expectation of a specific parameter: a probability, mean or median, respectively. But in addition, prediction markets also can be used to evaluate uncertainty about these expectations. For instance, consider a family of winner-take-all contracts that pay off if and only if the candidate earns 48 percent of the vote, 49 percent, 50 percent, and so on. This family of winner-take-all contracts will then reveal almost the entire probability distribution of the market's expectations. A family of spread betting contracts can yield similar insights. An even-money bet in a spread contract will define the median, as explained. But for similar reasons, a contract that costs \$4 and pays \$5 if $y>y^*$ will elicit a value of y^* that the market believes to be a four-fifths probability, thus identifying the 80th percentile of the distribution. As a final alternative, nonlinear index contracts can also reveal more information about the underlying distribution. For instance, consider a market with two index contracts, one that pays in a standard linear form and another that pays according to the square of the index, y^2. Market prices will reveal the market's expectation of $E[y^2]$ and $E[y]$, which can be used to make an inference about the market's beliefs regarding the standard deviation of $E[y]$, more commonly known as the standard error. (Recall that the standard deviation can be expressed as $\sqrt{(E[y^2] - E[y]^2)}$, or the square root of the mean of the squares less the square of the means.) By the same logic, adding even more complicated index contracts can yield insight into higher order moments of the distribution.

APPLICATIONS AND EVIDENCE

Perhaps the best-known prediction market among economists is the Iowa Electronic Market, run by the University of Iowa. The original Iowa experiment, run in 1988, allowed trade in a contract that would pay $2^1/_2$ cents for each percentage point of the popular vote in the presidential election won by Bush, Dukakis, or others. More recently, it has run markets based on the 2003 California gubernatorial election, the 2004 presidential election, the 2004 Democratic presidential nomination, and how the Federal Reserve will alter the federal funds interest rate. Universities in other countries have also started running event markets about their own elections, such as the Austrian Electronic Market run by the Vienna University of Technology or the University of British Columbia Election Stock Market that focuses on Canadian elections.

There are a growing number of Web-based event markets, often run by companies that provide a range of trading and gambling services. Some prominent examples include InTrade.com and Betfair.com and pseudomarkets (in which participants trade virtual currency), such as Newsfutures.com and Ideosphere.com. These Web sites often take the lead on defining a contract (as in the example of Poindexter's departure from DARPA described earlier), but then allow individuals to post their offers and to accept the offers of others.

Some prediction markets focus on economic statistics. The example of the Iowa market on the federal funds rate was mentioned earlier. More recently, Goldman Sachs and Deutsche Bank have launched markets on the likely outcome of future readings of economic statistics, including employment, retail sales, industrial production, and business confidence. The Chicago Mercantile Exchange is planning to open a market in inflation futures. Some event markets also forecast private sector returns. The Hollywood Stock Exchange allows people to use virtual currency to speculate on movie-related questions, such as opening-weekend performance, total box office returns, and who will win Oscars. In several cases, private firms have found innovative ways to use prediction markets as a business forecasting tool.

Exhibit 12.2 lists some of these prediction markets. Drawing on experiences with event markets, it is possible to start suggesting some generalizations about how prediction markets work, both in terms of their accuracy and whether arbitrage or market manipulation are possible.

ACCURACY OF PREDICTION MARKETS

Arguably the most important issue with these markets is their performance as predictive tools. In the political domain, Berg, Forsythe, Nelson, and Reitz (2001) summarize the evidence from the Iowa Electronic Markets, documenting that the market has both yielded very accurate predictions and also outperformed large-scale polling organizations. Exhibit 12.3 shows data from the past four U.S. presidential elections. The horizontal axis shows the number of days before the election. The vertical axis measures the average absolute deviation between the prices of index contracts linked to the two-party shares of the popular vote for each party and actual vote shares earned in the election. In the week leading up to the election, these markets have predicted vote shares for the Democratic and Republican candidates with an average absolute error of around 1.5 percentage points. By

Exhibit 12.2 Prediction Markets

Market	Focus	Typical Turnover on an Event ($US)
Iowa Electronic Markets: www.biz.iowa.edu/iem. Run by University of Iowa.	Small-scale election markets. Similar markets are run by: UBC (Canada): www.esm.buc.ca; and TUW (Austria): http://ebweb.tuwien.ac.at/apsm/.	Tens of thousands of dollars (Traders limited to $500 positions.)
Intrade: www.intrade.com. For-profit company.	Trade in a rich set of political futures, financial contracts, current events, and entertainment.	Hundreds of thousands of dollars.
Newsfutures: www.newsfutures.com. For-profit company.	Political, finance, current events, and sports markets. Also technology and pharmaceutical futures for specific clients.	Virtual currency redeemable for monthly prizes (e.g., a television).
Foresight Exchange: www.ideosphere.com. Nonprofit research group.	Political, finance, current events, science and technology events suggested by clients.	Virtual currency.
Hollywood Stock Exchange: www.hsx.com. Owned by Cantor Fitzgerald.	Success of movies, movie stars, awards, including a related set of complex derivatives and futures. Data used for market research.	Virtual currency.

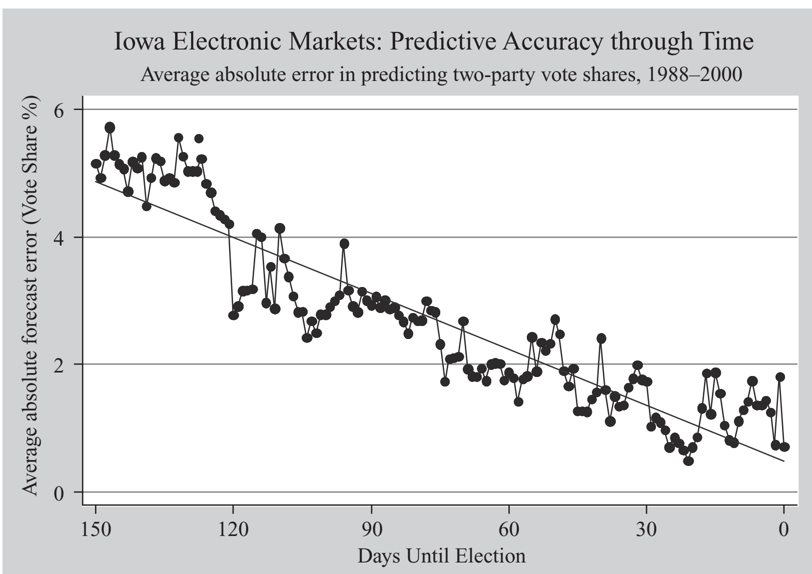


Exhibit 12.3 Information Revelation through Time

comparison, over the same four elections, the final Gallup poll yielded forecasts that erred by 2.1 percentage points. The graph also shows how the accuracy of the market prediction improve as information is revealed and absorbed as the election draws closer.

Perhaps more surprising in terms of how well prediction markets can aggregate information is the performance of markets at the level of the individual district. Typically districts are sufficiently small that there is little interest (or funding) for local polling, yet when Australian bookmakers started betting on district-level races, Leigh and Wolfers (2002) reported that they were extremely accurate.

That said, comparing the performance of markets with a mechanistic application of poll-based forecasting may not provide a particularly compelling comparison. A more relevant starting point might be to compare the predictions of markets with those of independent analysts. For an example along these lines, consider the Saddam Security, which was a contract offered on InTrade paying $100 if Saddam Hussein were ousted from power by the end of June 2003. Exhibit 12.4 shows that the price of this contract moved in lockstep with two other measures: expert opinion as shown by an expert journalist's estimate of the probability of the United States going to war with Iraq; and oil prices, an obvious barometer of political strife in the Middle East.

In a corporate context, the Hollywood Stock Exchange predicts opening weekend box office success, and Exhibit 12.5 shows that these predictions have been quite accurate. Further, this market has been about as accurate at forecasting Oscar winners as an expert panel (Pennock, Lawrence, Giles, and Nielsen, 2003). Some firms have also begun to experiment with internal prediction markets. An internal market at Hewlett-Packard (HP) produced more accurate forecasts of printer sales than the firm's internal processes (Chen and Plott, 2002). Ortner (1998) described an experiment at Siemens in which an internal market predicted that the firm would

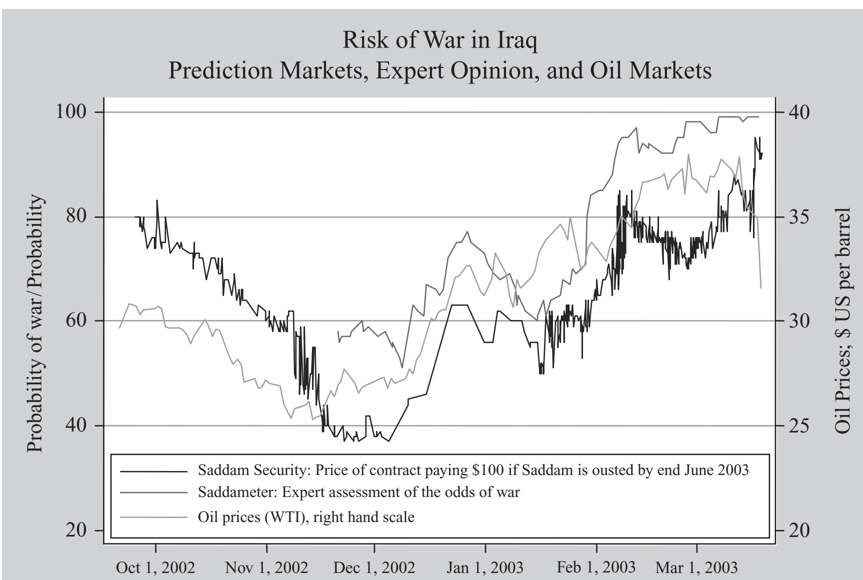


Exhibit 12.4 Saddam Security

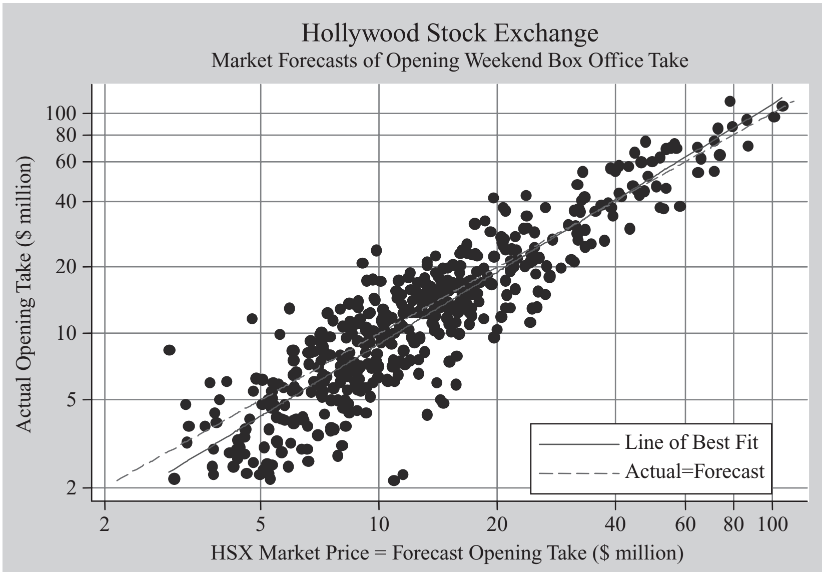


Exhibit 12.5 Predicting Movie Success

definitely fail to deliver on a software project on time, even when traditional planning tools suggested that the deadline could be met. While the Hollywood markets have drawn many participants simply on the basis of their entertainment value, the HP and Siemens experiences suggested that motivating employees to trade was a major challenge. In each case, the firms ran real money exchanges, with only a relatively small trading population (20–60 people), and subsidized participation in the market, by either endowing traders with a portfolio or matching initial deposits. The predictive performance of even these very thin markets was quite striking.

In another recent prediction market, traders in Economic Derivatives predict the likelihood that economic data released later in the week will take on specific values. The traditional approach to aggregating forecasts is simply to take an average or a "consensus estimate" from a survey of 50 or so professional forecasters. We now have data from the first year of operation of these markets. Exhibit 12.6 analyzes these early outcomes, comparing average market and consensus forecasts of three variables: total nonfarm payrolls data released by the Bureau of Labor Statistics; retail trade data (excluding autos) released by the Bureau of the Census; and business confidence as measured by the Institute for Supply Management's survey of manufacturing purchasing managers. The market-based predictions of these economic indicators are always extremely close to the corresponding "consensus" forecast, and hence, the two estimates are highly correlated. There are no statistically (or economically) meaningful differences in forecast performance—measured either as the correlation with actual outcomes or in terms of average absolute forecast errors. That said, this early sample is sufficiently small that precise conclusions are difficult to draw.

Interestingly, these markets yield not just a point estimate for each economic indicator but involve a menu of 10 to 20 winner-take-all contracts as to whether the

Exhibit 12.6 Predicting Economic Outcomes: Comparing Market-Aggregated Forecasts with Consensus Surveys

	Non-Farm Payrolls (Monthly change, '000s)	Retail Trade (ex autos) (Monthly change, %)	ISM Manufacturing Purchasing Managers' Index
Panel A: Correlations			
Corr (Market, Consensus)	0.91	0.94	0.95
Corr (Consensus, Actual)	0.26	0.70	0.83
Corr (Market, Actual)	0.22	0.73	0.91
Panel B: Mean absolute error			
Consensus	71.1	0.45	1.10
Market (empirical)	72.2	0.46	1.07
Market (implied expectation)	65.7	0.34	1.58
Panel C: Standard deviation of forecast errors (Standard error of forecast)			
Consensus	99.2	0.55	1.12
Market (empirical)	97.3	0.58	1.20
Market (implied expectation)	81.1	0.42	1.96
Sample size	16	12	11

Notes:
"Market" = market-implied mean forecast from www.economicderivatives.com.
"Consensus" = average of around 50 forecasters from www.briefing.com.
"Actual" = Preliminary estimates from original press releases (BLS, Census, ISM).

indicator will take on specific values. This family of contracts reveals an approximation to the full probability distribution of market expectations. Consequently, we can calculate the level of uncertainty surrounding specific point estimates. One measure of uncertainty is the expected absolute forecast error (although calculations using standard deviation provide the same qualitative results). The market-based assessments of uncertainty are shown in the last line of panel B. Comparing these implied expectations with outcomes in the first two rows of panel B suggests that the market-based assessments of uncertainty are of about the right magnitude. Finally, we can compare the implied standard errors of the forecasts with the reported standard errors of the statistics that the market is attempting to forecast. For instance, the Census Bureau reports that the change in retail trade is estimated with a standard error of around 0.5 percent, while the standard error implied by the prediction market is 0.42 percent. Taken literally, this suggests that the market believes that it is less uncertain about the Census Bureau estimate than the Census Bureau is.[4] Such results suggest either that the statistical agencies' errors are predictable, that their standard error estimates are (slightly) upwardly biased, or that traders are overconfident.

POSSIBILITIES FOR ARBITRAGE

Prediction markets appear to present few opportunities for arbitrage. There are several ways of looking for arbitrage opportunities: whether prices for similar contracts can be arbitraged across different exchanges or different securities; whether predictable patterns in the movement of the prices allow for arbitrage; and whether arbitrageurs might be able to exploit predictable deviations from rationality.

Exhibit 12.7 shows the bid and ask prices on a contract that paid \$100 if Schwarzenegger was elected California's governor in 2003, sampling data on bid and ask prices from two online exchanges every four hours. While both sets of data show substantial variation, they comove very closely, and opportunities for arbitrage (when the bid price on one exchange is higher than the ask on another) are virtually absent.

The pricing of families of related securities tends to be internally consistent. For example, Exhibit 12.8 shows the prices of several securities launched by InTrade that paid off if weapons of mass destruction were found in Iraq by May, June, July, or September 2003. Their prices moved closely together in a way that suggested that the prices of each contract digested similar information at close to the same time.

In most cases, the time series of prices in these markets does not appear to follow a predictable path, and simple betting strategies based on past prices appear to yield no profit opportunities; for example, Leigh, Wolfers, and Zitzewitz (2003) demonstrate this point for the InTrade Saddam Security. However, there is also some evidence that this small-scale market responded to news about Iraq with a slight lag relative to deeper financial markets. Tetlock (2004) surveys a wide range of data from InTrade, finding that their financial contracts are largely efficiently priced.

Prediction markets do seem to display some of the deviations from perfect rationality that appear in other financial markets. There is substantial evidence from psychology and economics suggesting that people tend to overvalue small

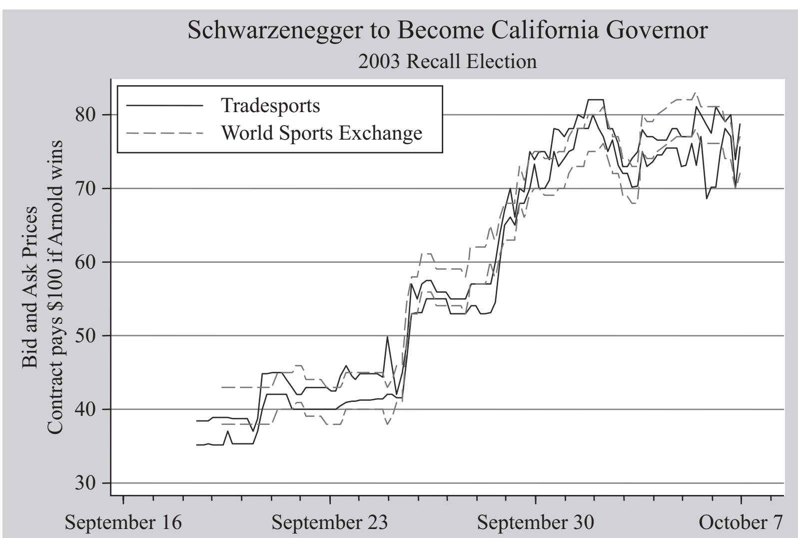


Exhibit 12.7 2003 California Gubernatorial Election

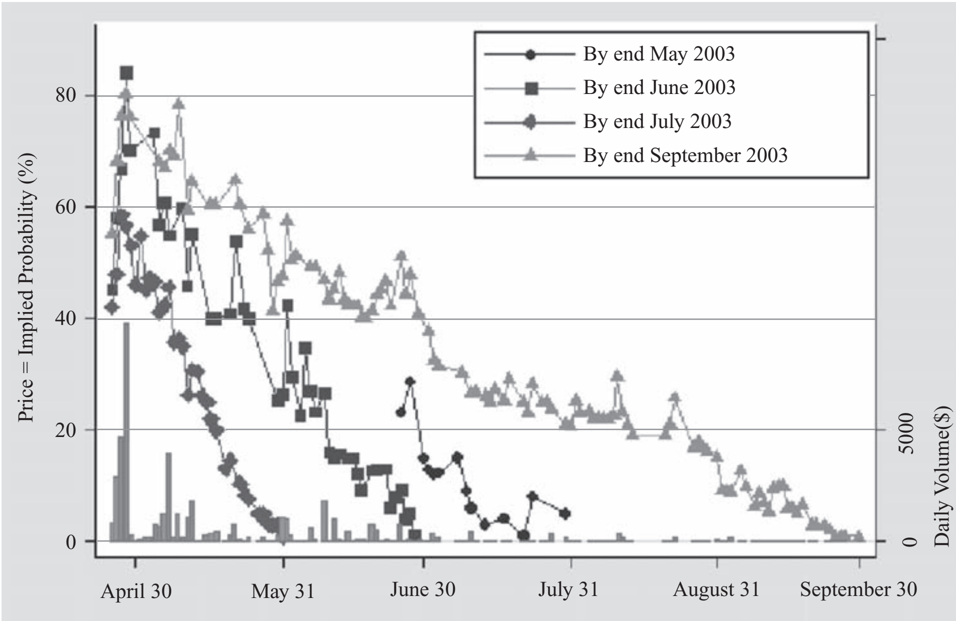


Exhibit 12.8 Will Weapons of Mass Destruction Be Discovered in Iraq?

probabilities and undervalue near certainties. For example, there is a well-known favorite/long-shot bias in horse races (discussed by Thaler and Ziemba, 1988), in which bettors tend to overvalue extreme long shots and thus receive much lower returns for such bets, an effect that is offset by somewhat higher (albeit still negative) returns for betting on favorites. The "volatility smile" in options refers to a related pattern in financial markets (Bates, 1991; Rubenstein, 1994), which involves overpricing of strongly out-of-the-money options and underpricing of strongly in-the-money options (relative to their future values or their ex-ante values from the Black-Scholes option pricing formula).[5] These experiences suggest that prediction markets may perform poorly at predicting small probability events.

An example of this kind of miscalibration comes from financial variables that trade on InTrade. Exhibit 12.9 reports the bid and ask prices in the prediction market for a contract that will pay $100 if the Standard & Poor's (S&P) 500 index finishes 2003 in a certain range. For comparison, one can look at the actual prices of December S&P options traded on the Chicago Mercantile Exchange. We used the method discussed in Leigh, Wolfers, and Zitzewitz (2003) to translate the financial market prices into prices for a security comparable to the InTrade contract. Comparing InTrade prices with the actual option prices in Chicago suggests that the extremely unlikely (high and low) outcomes for the S&P 500 are relatively overpriced on InTrade. In fact, the price differences implied a (small) arbitrage opportunity that persisted for most of summer 2003 and reappeared in 2004. Similar patterns existed for InTrade securities on other financial variables like crude oil, gold prices, and exchange rates. This finding is consistent with the long-shot bias being more pronounced on smaller-scale exchanges.

Exhibit 12.9 Price of Standard & Poor's Future Price Securities on InTrade versus Actual Prices from the Chicago Mercantile Exchange (Market close, July 23, 2003)

	Price on InTrade		
S&P Level at End of 2003	Bid	Ask	Estimated Price from Actual December S&P Options
1200 and over	2	6	2.5
1100 to 1199	11	16	13.2
1000 to 1099	28	33	33.3
900 to 999	25	30	30.5
800 to 899	14	19	13
700 to 799	3	8	5
600 to 699	4	7	2
Under 600	5	8	1
S&P level on July 23, 2003	985		

Notes: Prices given in first two columns are for a security that pays $100 if S&P finishes 2003 in given range. Prices in third column are estimated from actual option settlement prices using the method in Leigh et al. (2003), adjusting for the 13-day difference in expiry date.

Another behavioral bias reflects the tendency of market participants to trade according to their desires rather than objective probability assessments. Rhode and Strumpf (2004) provides evidence that certain New York gamblers are more likely to bet the Yankees, while Forsythe, Reitz, and Ross (1999) provide evidence that individual traders buy and sell in political markets in a manner correlated with their party identification. Even so, as long as marginal trades are motivated by profits rather than partisanship, prices will reflect the assessments of (unbiased) profit motive. Thus far, there is little evidence that these factors yield systematic unexploited profits.

A further possible limitation of prediction market pricing arises if speculative bubbles drive prices away from likely outcomes. Traditional markets may be subject to bubbles because of constraints on short selling and because investors will be reluctant to commit a large share of their wealth to an arbitrage opportunity, since if the mispricing does exist, it may be worse before it gets better (Shleifer and Vishny, 1997). Since prediction markets typically impose no restrictions on short selling, and the markets are sufficiently small scale that it is unlikely that informed investors will be capital constrained, the scope for bubbles might be more limited. It is impossible to make any serious attempt at describing the frequency of bubbles in the data we have so far. However, through September 2003, we suspected a bubble in the InTrade security on whether Hillary Clinton would win the Democratic nomination. Our suspicions were based on her public statements that she was not a candidate and the tenor of discussion among traders, which seemed to indicate that trading was being driven by expectations of future price movements rather than by fundamentals. Equally, these high prices may have reflected those with access to campaign insiders who knew more about her state of mind than we did.

Empirically, the best that we can say is that the performance of past markets at predicting the future has been, on average, pretty good, whether or not specific markets were in some cases distorted by biases or bubbles. Laboratory experiments hold out the possibility of learning more about bubbles, as it is possible for the experimenter to know the "true price" and, hence, to observe deviations. Plott and Sunder (1982, 1988) have set up extremely simple examples in which bubble-like behavior occurs in simple prediction markets. At the same time, bubbles in experimental markets often burst and give way to more rational pricing.

CAN EVENT MARKETS BE MANIPULATED EASILY?

The profit motive usually has proven sufficient to ensure that attempts at manipulating these markets were unsuccessful. There have been several known attempts at manipulation of these markets, but none of them had much of an effect on prices, except during a short transition phase.[6] For example, Wolfers and Leigh (2002) report candidates betting on themselves at long odds to create a buzz, while Strumpf (2004) placed random $500 bets on the Iowa Electronic Markets to trace their effect. In a similar vein, Camerer (1998) attempted with little effect to manipulate betting on horse races by canceling large wagers at the last moment. Clearly, the extent to which markets are manipulable depends, at least in part, on how thin the markets are.

It was feared that the DARPA markets would create the opportunity for a terrorist to profit from an act of terrorism or an assassination. With respect to the DARPA markets, this concern may have been misplaced, both because the proposed markets were unlikely to have included terrorism or assassination contracts in the first place and because the small scale of these markets means that terrorists would not have been able to earn much relative to the presumed going rate for an assassination. An alternative view holds that such trade is actually a good thing to the extent that trading ultimately reveals previously secret information about the intentions of terrorist groups. That said, if terrorists are sophisticated enough to place bets in futures markets, surely they can do so with standard futures contracts on oil prices, by selling short stock in insurance companies or the entire stock market and the like. Indeed, rumors have circulated widely that there was unusual trading in options on United and American Airlines stock in the week prior to the attacks of September 11, 2001. A careful analysis by Poteshman (2006) found little evidence to support these rumors, suggesting that if terrorists did profit from their actions, they neither left a noticeable footprint nor needed a prediction market to do so.

MARKET DESIGN

The success of prediction markets, like any market, can depend on their design and implementation. Some of the key design issues include how buyers are matched to sellers, the specification of the contract, whether real money is used, and whether a diversity of information exists in a way that provides a basis for trading. We consider these issues in turn.

In most prediction markets, the mechanism that matches buyers to sellers is a continuous double auction, with buyers submitting bids and sellers submitting asking prices and with the mechanism executing a trade whenever the two sides of the market reach a mutually agreeable price. However, the new prediction markets in announcements of economic statistics operate more like the parimutuel systems that are common in horse-race betting. In a parimutuel system, all of the money that is bet goes into a common pot and is then divided among the winners (after subtracting transaction costs). Many prediction markets are also augmented by market makers who announce willingness to buy and sell at a certain range of prices; similarly, most sports bets are placed with bookmakers who post prices. Finally, while these mechanisms are relatively useful for simple markets, Hanson (2003) has proposed the use of market scoring rules to allow for simultaneous predictions over many combinations of outcomes. Instead of requiring separate markets for each combination of possible outcomes, traders effectively bet that the sum of their errors over all predictions will be lower.

For a prediction market to work well, contracts must be clear, easily understood, and easily adjudicated. For example, we do not see contracts such as "Weapons of Mass Destruction Are Not in Iraq," but rather contracts specifying whether such weapons will have been found by a certain date. This requirement for clarity sometimes can turn out to be complex. In the 1994 U.S. Senate elections, the Iowa markets proposed what looked to be a well-specified market, with contracts paying according to the number of seats won by each party. The day after the election (and while votes were still being counted in some jurisdictions), Senator Richard Shelby (D-Alabama) switched sides to become a Republican. As another example, in the course of Ortner's (1998) internal prediction market on whether a software project would be delivered to the client on schedule, the client changed the deadline.

One intriguing question is how much difference it makes whether prediction markets are run with real money or with some form of play money. Legal restrictions on gambling have led some groups to adopt play money exchanges, with those who amass the largest play fortunes eligible for prizes. Prices on play and real-money exchanges are not linked by arbitrage: In August 2003, for example, George W. Bush was a 67 percent favorite to win reelection on real-money exchanges but was a 50-50 bet on NewsFutures. However, we do not yet have sufficient comparative data to know the extent to which money makes predictions more accurate. Indeed, it has been argued that the play money exchanges may outperform real-money exchanges because "wealth" can be accumulated only through a history of accurate prediction. In a suggestive experiment, Servan-Schreiber, Wolfers, Pennock, and Galebach (2004) compared the predictive power of the prices from a real-money and play-money exchanges over the 2003 NFL football season, finding that both yielded predictions that were approximately equally accurate. Interestingly, both sets of prices also outperformed all but a dozen of 3,000 people in an online contest and also easily outperformed the average assessments of these "experts." One practical advantage of play money contracts is that they offer more freedom to experiment with different kinds of contracts. On play money exchanges, such as Foresight Exchange, one often sees quite loosely worded "contracts" such as that a "scientific study will conclude that astrology is a statistically significant predictive method to describe an individual's personality traits."

Even well-designed markets will fail unless a motivation to trade exists.[7] Most prediction markets are not large enough to allow hedging against specific risks. However, the play money exchanges and sports gambling industry both suggest that it may be possible to motivate (small-scale) trading simply through the thrill of pitting one's judgment against others, and being able to win a monetary prize may sharpen this motivation. Trade also requires some disagreement about likely outcomes. Disagreement is unlikely among fully rational traders with common priors. It is more likely when traders are overconfident in the quality of their private information or their ability to process public information or when they have priors that are sufficiently different to allow them to agree to disagree.

These insights suggest that some prediction markets will work better when they concern events that are widely discussed, since trading on such events will have higher entertainment value and there will be more information on whose interpretation traders can disagree. Ambiguous public information may be better in motivating trade than private information, especially if the private information is concentrated, since a cadre of highly informed traders can easily drive out the partly informed, repressing trade to the point that the market barely exists. Indeed, attempts to set up markets on topics where insiders are likely to possess substantial information advantages typically have failed. For instance, the InTrade contracts on the next Supreme Court retirement or the future of the papacy have generated very little trade despite the inherent interest in these questions. Trade also can be subsidized either directly or indirectly by adding noise trades into the market, which provides the potential to profit from trading.

Finally, the power of prediction markets derives from the fact that they provide incentives for *truthful revelation*, they provide incentives for research and *information discovery*, and the market provides an algorithm for *aggregating opinions*. As such, these markets are unlikely to perform well when there is little useful intelligence to aggregate or when public information is selective, inaccurate, or misleading. Further, the weights that markets give to different opinions may not be an improvement on alternative algorithms where the accuracy of pundits is directly observable. For example, the public information on the probability of weapons of mass destruction in Iraq appears to have been of dubious quality, so it is perhaps unsurprising that both the markets were as susceptible as general public opinion to being misled.

MAKING INFERENCES FROM PREDICTION MARKETS

How might economists use the results from prediction markets in subsequent analysis? The most direct form of inference involves using these predictions directly. For instance, in their experiments at HP, Chen and Plott (2002) elicited expectations of future printer sales, which were of direct interest for internal planning purposes.

Some analyses have tried to link the time series of expectations elicited in prediction markets with time series of other variables. For instance, in Leigh et al. (2003), we interpreted movements in the Saddam Security as an index for the risk of war and interpreted the comovement with the oil price shown in Exhibit 12.4 as a causal relationship, concluding that war led to a $10 per barrel increase in oil prices. A similar analysis suggested that equity prices had built in a 15 percent war discount. Applying a similar methodology, Slemrod and Greimel (1999) linked the price of a Steve Forbes security in the 1996 Republican primary market with a rising interest rate premium on municipal bond prices, because Forbes's signature issue was a "flat tax" that would have eliminated the tax exemption for municipal bond interest. As with any regression context, one must be cautious before inferring that these correlations reflect causation and consider the issues of reverse causation, omitted variables, statistical significance, functional form, and the like.

It seems quite possible to design prediction market contracts so that they would bring out the connection between an event and other variables. For instance, in 2002, we could have floated two securities, one paying $\$P$ if Saddam were ousted in a year, where P is the future oil price, with the purchase price refunded otherwise, and another that paid $\$P$ if Saddam remains in power, again refunding the purchase price. The difference in the equilibrium price of these two securities can be interpreted as the market's expectation of the effect of ousting Saddam on oil prices. This inference does not require researchers to wait until sufficient variation in the political situation has accrued for a regression to be estimated. Moreover, changes in the market's beliefs about how ousting Saddam would affect oil prices can be directly measured through such a conditional market.

Very few of these contingent markets have been constructed, although the Iowa Electronic Market on the 2004 presidential election is instructive. Exhibit 12.10 shows the prices of a series of contracts that are standard index contracts that pay a penny for each percentage of the two-party popular vote won by each party but are contingent in that the contract pays out only if the Democratic nominee is also successfully predicted. These contracts pay nothing if the nominee is not correctly predicted.

Because the Democratic and Republican shares of the two-party vote must sum to 1, a portfolio containing contracts tied to both the Democratic and Republican vote shares, but conditional on Kerry winning the nomination, will definitely pay \$1 if Kerry wins the primary and \$0 otherwise. Implicitly, then, this market embeds a winner-take-all market on the Democratic primary race, and adding the prices shown in columns A and B yields the prices of these synthetic securities that

Exhibit 12.10 Contingent Markets: 2004 Presidential Election (Contracts Pay According to Vote Share, Conditional on the Democratic Nominee)

Contract Pays Conditional On Specific Democratic Candidate	Democratic Candidate Vote Share (Contract price, \$) A	Republican Vote Share against This Candidate (Contract price, \$) B	Implied Probability This Candidate Wins Nomination C = A + B	Expected Share of Popular Vote if Nominated D = A/C
John Kerry	\$0.344	\$0.342	68.6%	50.1%
John Edwards	\$0.082	\$0.066	14.8%	55.4%
Howard Dean	\$0.040	\$0.047	8.7%	46.0%
Wesley Clark	\$0.021	\$0.025	4.6%	45.7%
Other Democrats	\$0.015	\$0.017	3.2%	46.9%

Notes: Columns A and B show the prices of contracts that pay a penny for each percentage of the two-party popular vote won by Democrats or Republicans respectively, conditional on picking the winner of the Democratic nomination. (Contracts pay \$0 if the selected candidate does not win the Democratic nomination.)

Source: Closing prices January 29, 2004, Iowa electronic markets.

represent the probability that any specific candidate wins the Democratic nomination (shown in column C). The final column calculates the implied expected vote share for each candidate, if that candidate were to win the nomination, by deflating the cost of the Democratic vote share contract conditional on that candidate by the probability of that candidate actually winning the nomination. Hanson (1999) has called these contingent markets "decision markets," arguing that these expectations should be used to guide decision making. As such, delegates to the Democratic convention interested in selecting the strongest candidate would simply compare the ratios in the final column, and accordingly, vote for John Edwards. Berg and Rietz (2003) make a related argument using data from the 1996 Republican nomination race.

While we are optimistic that these data on contingent prediction markets can be used to inform decision making, some care is required. In making statements about the comovement of two variables, social scientists have long struggled to distinguish correlation from causation, and these decision markets do not resolve this issue. One could imagine that traders hold a frequentist view of probability and that they price the securities in Exhibit 12.10 by simply inventing hundreds of possible scenarios, and prices simply reflect average outcomes across these scenarios. An econometrician running regressions based on these hundreds of scenarios would note a robust correlation between Edwards winning the nomination and the Democrats winning the presidency. But a careful econometrician would be reluctant to infer causation, noting that there are important "selection effects" at play, as the scenarios in which Edwards wins the nomination are not random. For example, the markets may believe that Edwards will not win the nomination unless Southern Democrats become energized, but if this does happen, it is likely that Edwards will win both the nomination and the presidency. Alternatively, with Kerry viewed as the likely nominee, Edwards may be perceived as a possible nominee only if he shows himself to be a politician of extraordinary ability, overcoming Kerry's early lead in the delegate count. If so, it also seems likely that a candidate of such extraordinary ability would win the general election. Or Edwards might be perceived as thin-skinned and likely to drop out of the race if it appears that the Democratic party is unlikely to win the White House. As such, the relatively high price of the Edwards-Democratic security may reflect either something about Edwards's ability or the selection effects that lead him to win the nomination.

Just as econometricians often deal with selection effects by adding another equation that explicitly models the selection process, there is a prediction market analog—floating another contract that prices the variables driving the selection of Democratic candidates. For example, adding a contract that pays off if a candidate drops out of the nomination race early would allow an assessment of the extent to which prices of contingent contracts are being driven by that specific selection mechanism, thereby yielding a more accurate indication of candidate ability. But since many key traits of candidates may be unobservable, or difficult to capture in a contract that would attract trading, it may be impossible to rely fully on contingent markets to guide voters to the candidate with the greatest vote-winning potential.[8]

These relatively simple contingent markets, as well as more complex combinatorial markets, are as yet virtually untested and a useful focus for further research.

There may be important and interesting applications in domains where selection problems are minimal.

INNOVATIVE FUTURE APPLICATIONS?

Prediction markets are extremely useful for estimating the market's expectation of certain events; simple market designs can elicit expected means or probabilities, more complex markets can elicit variances, and contingent markets can be used to elicit the market's expectations of covariances and correlations, although as with any estimation context, further identifying assumptions are required before a causal interpretation can be made. The research agenda on these markets has reflected an interplay between theory, experiments, and field research, drawing on scholars from economics, finance, political science, psychology, and computer science. This research program has established that prediction markets provide three important roles:

1. Incentives to seek information.
2. Incentives for truthful information revelation.
3. An algorithm for aggregating diverse opinions.

Current research is only starting to disentangle the extent to which the remarkable predictive power of markets derives from each of these forces.

Prediction markets doubtless have their limitations, but they may be useful as a supplement to the other relatively primitive mechanisms for predicting the future, such as opinion surveys, politically appointed panels of experts, hiring consultants, or holding committee meetings. We are already seeing increasing interest in these markets in the private sector, with the experiments at Hewlett-Packard now being supplemented with new markets on pharmaceuticals and the likely success of future technologies on NewsFutures.

DARPA's ill-fated attempt at establishing a Policy Analysis Market ultimately failed. However, it seems likely that private sector firms will continue to innovate and to create new prediction markets, so policy makers will still be able to turn to prediction markets run by firms like InTrade, Net Exchange, Incentive Markets, and NewsFutures. It may be a sensible political outcome to have these event markets run by publicly regulated, private sector firms. Nonetheless, to the extent that the valuable information generated by trade in these markets is not fully internalized into the profits earned by these private firms, prediction markets may be underprovided.

ACKNOWLEDGMENTS

The authors would like to thank David Pennock, Emile Servan-Schreiber of NewsFutures, David Dempsey and John Delaney of InTrade, Alison Fealey and Oliver Frankel of Goldman Sachs, and George Neumann of IEM for help with data. Thanks to Kay-Yut Chen, Robin Hanson, Andrew Leigh, Betsey Stevenson, Timothy Taylor, Hal Varian, Craig Yee, and William Ziemba for stimulating discussions. Matthew Abreu, Rohak Doshi, Doug Geyser, Mike Goelzer, Chris Lion, Paul Reist, Erik Snowberg, and Ravi Pillai provided outstanding research assistance.

ENDNOTES

1. Looney (2003) provides a useful summary of both the relevant proposal and its aftermath. Further, Robin Hanson has maintained a useful archive of related news stories and government documents at http://hanson.gmu.edu/policyanalysismarket.html.
2. For a vision of how prediction markets, if they develop sufficient liquidity, may also prove useful for those wishing to hedge against specific risks, see the discussions in Athanasoulis, Shiller, and van Wincoop (1999) and Shiller (2003).
3. The price of a winner-take-all security is essentially a state price, which will equal an estimate of the event's probability under the assumption of risk neutrality. The sums wagered in prediction markets are typically small enough that assuming that investors are not averse to the idiosyncratic risk involved seems reasonable. But if the event in question is correlated with investors' marginal utility of wealth, then probabilities and state prices can differ. In what follows, we leave this issue aside and use the term probability to refer to risk-neutral probability.
4. A similar comparison can be made for nonfarm payrolls, although the inference is less direct. The U.S. Bureau of Labor Statistics estimates that their final estimate of the change in nonfarm payrolls has a standard error of around 64,000, while the preliminary estimate is more uncertain. The BLS has yet to estimate a standard error for their preliminary estimates, but the root mean squared error of the preliminary estimate relative to the final estimate is around 50,000. If the revision to the preliminary estimate and the subsequent error in the revised estimate were uncorrelated, this would imply a standard error for the preliminary estimate of about 81,500. Comparing these numbers with the average standard error of the market forecast of 81,100 suggests that the market is about as sure of the advance estimate as the BLS.
5. Aït-Sahalia, Wang, and Yared (2001) argue that the conclusion of miscalibration is less clear cut in this context, because these prices may be driven by small likelihoods of extreme price changes.
6. Rhode and Strumpf (2004) document that attempts at manipulation in early twentieth-century political markets were typically unsuccessful.
7. The inflation futures market on the Coffee, Sugar, and Cocoa Exchange is a case in point; this market generated little volume, ultimately failing.
8. Furthermore, it is worth noting that the incentives to manipulate a contract rise with its use in decision making, and the apparent failure of the past manipulation attempts mentioned above do not guarantee that it would fail in this context.

REFERENCES

Aït-Sahalia, Yacine, Yubo Wang, and Francis Yared. 2001. "Do Options Markets Correctly Price the Probabilities of Movement of the Underlying Asset?" *Journal of Econometrics* 102: 67–110.

Athanasoulis, Stefano, Robert Shiller, and Eric van Wincoop. 1999. "Macro markets and financial security," *Economic Policy Review* 21–39.

Bates, David. 1991. "The Crash of '87: Was It Expected? The Evidence from Options Markets," *Journal of Finance* 46(3): 1009–44.

Berg, Joyce, Robert Forsythe, Forrest Nelson and Thomas Rietz. 2001. "Results from a Dozen Years of Election Futures Markets Research," forthcoming in Charles Plott and Vernon Smith (Eds) *Handbook of Experimental Economic Results*.

Berg, Joyce and Thomas Rietz. 2003. "Prediction Markets as Decision Support Systems," *Information Systems Frontiers* 5(1).

Camerer, Colin. 1998. "Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting," *Journal of Political Economy* 106(3).

Chen, Kay-Yut and Charles Plott. 2002. "Information Aggregation Mechanisms: Concept, Design and Field Implementation for a Sales Forecasting Problem," *Working Papers 1131*, California Institute of Technology, Division of the Humanities and Social Sciences.

Forsythe, Robert, Thomas Rietz and Thomas Ross. 1999. "Wishes, Expectations and Actions: Price Formation in Election Stock Markets," *Journal of Economic Behavior and Organization* 39: 83–110.

Hanson, Robin. 1999. "Decision Markets," *IEEE Intelligent Systems* 14(3): 16–19.

Hanson, Robin. 2003. "Combinatorial Information Market Design," *Information Systems Frontiers* 5(1): 105–119.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica* 47(1): 153–61.

Leigh, Andrew and Justin Wolfers. 2002. "Three Tools for Forecasting Federal Elections: Lessons from 2001," *Australian Journal of Political Science* 37(2).

Leigh, Andrew, Justin Wolfers and Eric Zitzewitz. 2003. "What do Financial Markets Think of War in Iraq?," *NBER Working Paper 9587.*

Looney, Robert. 2003. "DARPA's Policy Analysis Market for Intelligence: Outside the Box or Off the Wall," *Strategic Insights*, Vol. II Issue 9, September 2003.

Ortner, Gerhard. 1998. "Forecasting Markets – An Industrial Application," *mimeo*, Technical University of Vienna.

Pennock, David, Steve Lawrence, C. Lee Giles, and Finn Arup Nielsen. 2001. "The Real Power of Artifical Markets," *Science* 291: 987–988.

Plott, Charles and Shyam Sunder. 1982. "Efficiency of experimental security markets with insider information: An application of rational-expectations models," *Journal of Political Economy* 90(4): 663–98.

Plott, Charles and Shyam Sunder. 1988. "Rational Expectations and The Aggregation of Diverse information in Laboratory Security Markets," *Econometrica* 56: 1085–1118.

Poteshman, Allen. 2006. "Unusual Option Market Activity and the Terrorist Attacks of September 11, 2001," *University of Chicago Press, Journal of Business* 79: 1703–1726.

Rhode, Paul and Koleman Strumpf. 2004. "Historical Presidential Betting Markets," *Journal of Economic Perspectives*, Spring 2004.

Rubenstein, Mark. 1994. "Implied Binomial Trees," *Journal of Finance* 49: 771–818.

Servan-Schreiber, Emile, Justin Wolfers, David Pennock and Brian Galebach. 2004. "Prediction Markets: Does Money Matter?," *Electronic Markets* 14: 3.

Shiller, Robert. 2003. *The New Financial Order: Risk in the Twenty-first Century*, Princeton University Press: Princeton, NJ.

Shleifer, Andrei and Robert Vishny. 1997. "The Limits of Arbitrage," *Journal of Finance* 52(1): 35–55.

Slemrod, Joel and Timoth Greimel. 1999. "Did Steve Forbes Scare the Municipal Bond Market?," *Journal of Public Economics* 74(1): 81–96.

Spann, Martin and Bernd Skiera. 2003. "Internet-Based Virtual Stock Markets for Business Forecasting," *Management Science* 49: 1310–1326.

Strumpf, Koleman and Paul Rhode. 2008. "Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data," *mimeo*, Kansas University.

Tetlock, Paul. 2004. "How Efficient are Information Markets? Evidence from an Online Exchange," *mimeo*, Columbia GSB.

Thaler, Richard H., and William T. Ziemba. 1988. "Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries," *Journal of Economic Perspectives*, Vol. 2, No. 2. Spring, pp. 161–174.

ABOUT THE AUTHORS

Justin Wolfers is an associate professor of economics in the Business and Public Policy Department at the Wharton School. He is a visiting scholar with the Philadelphia and San Francisco Federal Reserve Banks, a research fellow with the National Bureau of Economic Research, the Center for Economic Studies in Munich, and the Institute for the Study of Labor in Bonn, and a research affiliate with the Centre for Economic Policy Research in London. He was previously an assistant professor at Stanford's Graduate School of Business. He earned his Ph.D. in economics in 2001 from Harvard University, and was a Fulbright, Knox and Menzies Scholar. His research focuses on labor economics, macroeconomics, law and economics, social policy and behavioral economics. Beyond research, he is a popular MBA teacher, a contributor to the New York Times' Freakonomics blog, a Wall Street Journal columnist and is a frequent contributor to the public debate.

Eric Zitzewitz is an associate professor of economics at Dartmouth College. He earned an AB in economics from Harvard College in 1993 and a Ph.D. in economics from MIT in 2001 and has previously taught at Stanford and Columbia business schools. His research includes work on agency problems and conflicts of interest in financial settings and on forecasting and prediction markets and has been published in numerous leading journals.

CHAPTER 13

Credit Default Swaps

STEVEN TODD
Associate Professor of Finance, Loyola University, Chicago

Credit default swaps (CDS) are the building blocks for all credit derivatives, from index correlation trades to synthetic collateralized debt obligations (CDOs). A CDS is similar to an insurance contract between two parties, a protection seller (insurer) and a protection buyer (insured). The protection buyer seeks to protect an asset from a loss of principal. The protection seller agrees to provide insurance for a fee. We say that the protection seller is long the underlying credit risk and the protection buyer is short the underlying credit risk. The underlying asset, known as the reference entity or obligation, can be a bond or leveraged loan of a corporate issuer, a basket of bonds or loans, sovereign debt, an asset-backed security (ABS), or a tranche from a CDO.

The protection buyer makes periodic fixed payments to the protection seller, until the contract expires, or, in the case of corporate credit default swaps, until the earlier of contract expiration or a credit event (see Exhibit 13.1). When the underlying asset is corporate debt, credit events may include issuer bankruptcy, failure to pay principal and sometimes debt restructuring events (for European obligors and North American investment-grade entities). For structured finance assets and CDOs, the definition of a credit event is expanded to include interest shortfalls and principal write-downs or implied write-downs.

CDS contracts can be settled via physical delivery of the underlying asset or with a cash payment. With cash settlement, the protection seller pays the protection buyer a cash amount equal to the difference between the par amount of the underlying asset and its recovery value. With physical settlement, the protection buyer delivers the reference obligation to the protection seller in exchange for a par payment.

CREDIT DEFAULT SWAPS ON CORPORATE DEBT

Corporate debt credit default swaps first appeared in the mid-1990s. These were later followed by CDS on structured credit assets, including residential mortgage-backed securities (RMBS) and commercial mortgage-backed securities (CMBS). CDS on CDOs are the most recent innovation. According to the Bank for International Settlements (BIS), the market for CDS has experienced dramatic growth, from a notional amount of $180 billion in 1997 to more than $45 trillion as of June 2007. Most of the outstanding contracts reference corporate entities. The

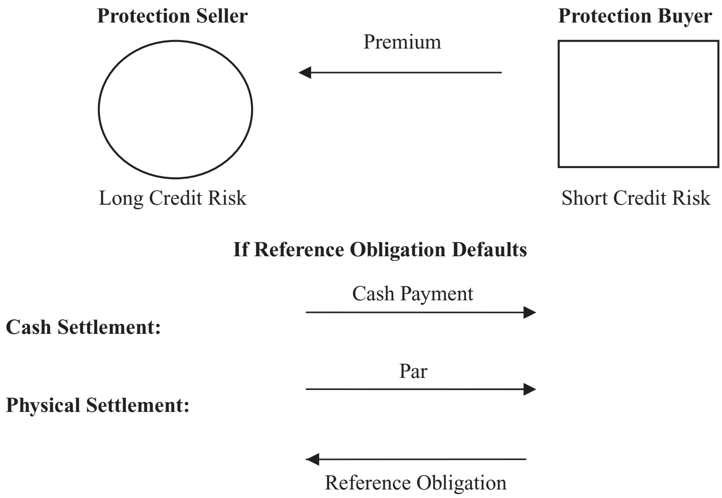


Exhibit 13.1 Credit Default Swap

International Swaps and Derivatives Association (ISDA), a trade organization of participants in the over-the-counter (OTC) derivatives market, has been influential in standardizing CDS contracts.

For CDS on corporate debt, credit events typically are limited to the bankruptcy of the referenced corporate entity and failure to pay principal on the debt. CDS written on European companies, sovereigns, and North American investment-grade corporate entities generally expand the list of credit events to include "restructuring" events. Restructuring events encompass efforts by corporations to preempt formal insolvency (bankruptcy) proceedings by negotiating changes in debt terms with creditors (see Exhibit 13.1).

Most corporate credit default swaps reference the senior, unsecured debt of a company. The most liquid contracts mature 3, 5, 7, or 10 years postsettlement, but any maturity is possible since the contracts are negotiated privately. If a credit event results in physical settlement, the protection buyer can deliver any senior, unsecured (unsubordinated) bond or loan that matures in 30 years or less and that is denominated in a standard currency. In general, protection buyers will favor the cheapest-to-deliver asset. Physical settlement may suffer from operational problems when the notional amount of synthetic assets exceeds that of eligible deliverable assets. For this reason, corporate credit default swaps allow for cash settlement, although cash settlement can present an alternative set of challenges when illiquid, distressed assets trade with wide bid/ask spreads.

CREDIT DEFAULT SWAPS ON ASSET-BACKED SECURITIES

Unlike corporate credit default swaps that can reference multiple loans or bonds of a corporation, ABS CDS reference specific tranches (notes) of asset-backed securities. These are structured investments collateralized by consumer assets, such as credit card receivables, mortgage loans, student loans, or auto loans and leases. The largest sector of the ABS market is the residential mortgage sector, known as the home equity loan (HEL) market.

The mortgage loans that comprise the HEL ABS sector generally are loans that do not meet the underwriting standards of the government-sponsored agencies (Government National Mortgage Association, Federal National Mortgage Association, and Federal Home Loan Mortgage Corporation), which place limits on loan size, loan type, borrower credit rating, and so on. These nonconforming loans include jumbo loans to prime borrowers (Alt-A loans) as well as loans to subprime and near-prime borrowers. Prime borrowers are those with FICO (Fair Isaac & Co.) scores above 679; subprime borrowers are those with FICO scores below 621.

Cash-out refinance loans, which allow borrowers to access the equity in their homes, dominate the collateral backing most residential ABS transactions. These loans may be fixed rate, adjustable rate, or interest-only (IO) loans. Adjustable rate mortgage loans have coupons that float above a benchmark interest rate, such as the London Interbank Offered Rate (LIBOR). An initial teaser rate may be used to provide lower monthly borrower payments initially. The time to first rate reset varies from 2 years (so called 2/28 loans) to 5 years (5/25 loans).

The loans that collateralize HEL ABS structures may be fixed rate only, adjustable rate only, or a combination of fixed and adjustable rate loans. These assets can support notes (liabilities) that are fixed rate, adjustable rate, or a combination of fixed and adjustable rates.

The first credit default swaps on ABS tranches followed the cash/physical settlement protocol of corporate CDS, terminating with a credit event payment from the protection seller to the protection buyer. This proved unsatisfactory because the principal and interest payments on asset-backed securities do not resemble those of corporate bonds. Moreover, default is a more elusive concept for structured finance assets, where principal write-downs may be reversed with subsequent write-ups, and interest shortfalls can be accrued and paid out later.

The International Swaps and Derivatives Association (ISDA) acknowledged these issues and introduced a pay-as-you-go (PAYGO or PAUG) template for credit default swaps on asset-backed securities in June 2005. The collapse of the subprime loan market in 2007 and subsequent initiatives by policy makers to stem the crisis have led to several legal tests of ISDA's template. Instead of a single credit event triggering a payment, PAUG provides for two-way payments between protection buyers and protection sellers. Each period, a protection buyer (the fixed rate payer) pays a protection seller (the floating rate payer) a fixed payment equal to the product of the fixed rate (premium), the average notional balance of the credit default swap, and the actual number of days in the calculation period, divided by 360. The notional value of the credit default swap tracks the par value of the reference entity, which adjusts for principal payments, prepayments, and write-downs.

The PAUG template allows for multiple trigger events, called floating amount events. These include interest and principal shortfalls and principal write-downs. Floating amount events trigger a floating rate payment from the protection seller to the protection buyer. If a shortfall or write-down is later reversed, the protection buyer repays the protection seller an additional fixed payment equal to the shortfall or write-up amount.

Interest shortfalls are reimbursed one of three ways, depending on which interest rate standard or toggle feature the CDS contract employs. With the fixed

Exhibit 13.2 Toggle Features

		Protection Seller
	Protection Seller Pays	**Net Maximum Payment**
Fixed Cap	Fixed premium	0
Variable Cap	LIBOR + Fixed premium	LIBOR
No Cap	Full amount of interest Shortfall	LIBOR + Spread on Reference Obligation – Fixed Premium

cap option, interest shortfalls are limited to an amount equal to the fixed rate payment, the periodic payment that the protection buyer pays to the protection seller. In contrast, with the no-cap option, the protection seller covers the full amount of any interest shortfall, even if it exceeds the fixed payment. The variable cap standard lies in the middle of these two options, limiting the protection seller's liability to LIBOR plus the fixed payment amount (see Exhibit 13.2). Protection buyers enjoy the greatest protection with the no-cap option; in contrast, protection sellers prefer to write contracts with the fixed cap option.

In contrast to the floating-amount events just described, credit events include failure to pay principal, principal write-downs, and a downgrade of the underlying asset to Caa2/CCC or below. If a credit event occurs, the protection buyer has the option to continue the contract or demand full or partial settlement. With partial settlement, the PAUG contract remains in force for the remaining notional amount. CDS contracts on ABS tranches typically have long maturities, matching those of the reference entities.

CREDIT DEFAULT SWAPS ON COLLATERALIZED DEBT OBLIGATIONS

The first collateralized debt obligation (CDO) was structured in 1988 and was backed by a collateral pool consisting of high-yield bonds—a collateralized bond obligation, or CBO. Since then, issuers have expanded the menu of reference assets to include leveraged loans (broadly syndicated and middle-market loans), trust preferred securities, emerging market sovereign debt, asset-backed securities, commercial mortgage-backed securities, commodities, and municipal debt.[1] CDO technology is adaptable, and increasingly issuers are using synthetic assets (loans or securities referenced via credit default swaps) to collateralize an issue.

A CDO is sliced into several notes (or tranches) that represent the issue's liabilities. A note's position in this capital structure determines its priority of claims on the collateral cash flows, which in turn determines its rating. Ratings can range from triple-A for the senior classes to unrated for the first-loss (or equity) tranche. A CDO only redistributes among the newly created securities the total credit risk associated with the pool of assets.

The template for CDO credit default swaps resembles the standardized template for ABS CDS. However, adjustments are made to accommodate implied principal write-downs and payment-in-kind (PIK) events. Asset-backed securities respond to severe collateral stresses by reducing (writing down) the par value of junior notes. In contrast, CDOs use a waterfall mechanism to reallocate cash flows to senior notes.[2] If an overcollateralization (O/C) test fails, the scheduled interest payments for a junior note can be diverted to pay down the principal of a more senior note. When this occurs, we say that the subordinate note PIKs. In general, some junior CDO notes may PIK until O/C compliance is achieved. With severe collateral stresses, however, O/C failures may persist and implied principal write-downs become inevitable. Because of implied principal write-downs and PIK events, a CDO credit default swap may fail to replicate the cash flows of its underlying reference asset.

In Exhibit 13.3, we compare the key features of ABS and CDO CDS. Both templates reference a specific tranche or obligation, and both use pay-as-you-go settlement. Buyers of protection pay sellers of protection an annual amount equal to the product of the notional principal and the protection premium (in quarterly or monthly installments). The notional principal value declines as the reference asset amortizes. A credit event occurs if the reference asset experiences a missed payment (principal or interest) or if it is downgraded to a distressed rating. For PIK-able CDO assets, the interest shortfall must persist for 360 consecutive days for a credit event to be deemed to have occurred. If a credit event occurs, the protection buyer can deliver the asset to the protection seller in exchange for a par value payment.

Less severe floating amount events occur if there is an interest shortfall or principal write-down. ABS credit default swap contracts list both actual and implied write-downs as credit/floating payment events. In contrast, CDO credit default swaps employ a toggle option for implied write-downs based on O/C ratios. Generally, protection sellers compensate protection buyers when an O/C ratio falls below 100 percent. Protection buyers find the implied write-down toggle appealing because it capitalizes the losses from a deteriorating collateral pool early rather than at the note's legal final maturity.

Exhibit 13.3 Credit Default Swaps on ABS and CDOs

	CDS on ABS	**CDS on CDOs**
Reference Obligation	ABS Bond – CUSIP specific	CDO Tranche – CUSIP specific
Settlement	Physical, PAUG	Physical, PAUG
Deliverable	ABS Bond – CUSIP specific	CDO Tranche – CUSIP specific
Maturity	Same as reference obligation	Same as reference obligation
One-time Credit Events	Failure to pay principal by maturity Downgrade to CCC/Caa2 or below Write-down	Failure to pay principal by maturity Downgrade to CCC/Caa2 or below Failure to pay interest on non-PIK-able Continuing failure to pay interest
Reversible Floating Events	Failure to pay principal	Failure to pay principal

Interest shortfalls lead to different cash flows under three possible cap options. In general, protection buyers (who are often hedging long positions in ABS or CDO assets) enjoy the most coverage with a no-cap option. Conversely, the obligations of protection sellers are minimized with a fixed-cap option. The fixed cap basis is usually paired with the implied write-down toggle for credit default swaps referencing senior notes, which are not usually PIK-able. For CDS referencing mezzanine tranches (which are typically PIK-able), a variable cap basis is usually paired with a no–implied–write-down toggle.

THE BASIS

In theory, the spread on a CDS contract should closely track the effective spread (after price discounts or premiums) on the underlying cash asset. In reality, CDS and cash spreads often diverge. The difference between the CDS premium and the underlying reference entity's asset-swap spread defines the basis. When the CDS spread is less (greater) than the asset-swap spread, the basis is said to be negative (positive). A negative or positive basis is possible if the risk profile of a CDS contract diverges from that of the cash asset. The basis is also influenced by a host of fundamental and technical factors, such as the supply of and demand for liquidity, financing costs, hedging activities, and dealer pricing power.

Because many cash investors are buy-and-hold accounts and CDS contracts facilitate leveraged short (and long) credit positions, the synthetic markets serve as marginal providers of liquidity. As a result, synthetic spreads generally lead cash spreads.

Theoretically, a highly negative basis should encourage negative basis trades. Here an investor takes a long credit position in the cash market while simultaneously buying protection in the synthetic market. However, a clean swap occurs only if the CDS contract perfectly mimics the cash asset. The pay-as-you-go protocol of ABS CDS likely produces a good match. In contrast, corporate CDS often introduce basis risk because the cash asset may be called or the underlying CDS could be canceled in response to a corporate restructuring event.

CDS INDICES

A CDS index is a basket of single-name credit default swaps, usually equally weighted. The basket may be formed by combining credit default swaps on corporate bonds, structured product securities, or leveraged loans. Currently, there is no CDS index of CDOs. CDS indices incorporate the same mechanics as single-name credit default swaps. A protection buyer makes periodic payments to a protection seller and the contract remains in effect until maturity (usually between 3 and 10 years) or, until a credit event occurs, in the case of corporate indices.

With the corporate indices, when a credit event occurs, the protection buyer has the right to sell the defaulted security at par to the protection seller. The defaulted asset is then removed from the index and the contract continues with a reduced notional for its remaining term. The asset-backed indices do not allow for physical settlement and instead use modified pay-as-you-go settlement rules that closely mirror standardized ABS CDS templates. Generally, the notional amount of an asset-backed security index adjusts as the reference entities amortize, prepay, default, or incur write-downs or write-ups. Index attributes for the most actively traded indices are summarized in Exhibit 13.4.

CDS indexes are static, which means the assets that populate a particular series are fixed; absent any credit events, the contents of the basket do not change over time. In contrast, most CDO portfolios are dynamic. Subject to various trading restrictions, a CDO manager may buy and sell assets as the contents of the portfolio amortize, prepay, or mature.

The contents of credit default swap indices are determined by polling a global group of broker-dealers, who serve as market makers. Every six months an index "rolls." This means that a new "on-the-run" series is created from the most liquid single-name CDS contracts. Since dealers have a vested interest in maintaining some degree of continuity among the series, many reference names remain unchanged when an index rolls. However, all defaulted securities and some downgraded credits (fallen angels) are replaced by new names.

CDS indices trade in both funded and unfunded forms. In an unfunded trade, the protection seller makes no down payment, exposing the protection buyer to counterparty risk. A fully funded trade eliminates all counterparty risk by requiring the protection seller to invest the trade notional amount in a portfolio of low-risk (typically AAA) securities. The collapse of the subprime loan market in 2007 and the ensuing credit crisis alerted many investors to the significance of counterparty risks. In response, a consortium of CDS broker-dealers have proposed launching a clearinghouse or exchange to insure against counterparty failures. The first CDS indices that were created were corporate indices, launched in 2003. In 2004, they merged to form a new set of indices called iTraxx (Europe, Australia, and Asia) and CDX (North American and Emerging Markets). These indices are divided into subindices, based on geography, rating, and sector. So, for example, iTraxx Europe TMT specializes in European investment-grade entities from the telecom, media, and technology sectors. Similarly, CDX.NA.IG specializes in North American investment-grade entities.

The U.S. and European CDS indices define credit events differently. For the U.S. indices, only bankruptcy and failure to pay trigger a default. Restructuring is not deemed a credit event, even though most underlying single-name CDS contracts treat restructuring as a credit event. European indices trade with the same credit events as the underlying CDS contracts, including modified restructuring, bankruptcy, and failure to pay.

The corporate indices have the option of physical settlement or cash settlement based on an auction price. If the notional amount of synthetic assets exceeds the par value of deliverable assets, physical settlement may suffer from operational problems. To avoid a short squeeze, corporate CDS index contracts allow for cash settlement. Cash settlement may present additional challenges if assets are illiquid (trading with wide bid-ask spreads) or distressed. To reduce price uncertainty, Markit—a firm that provides information about and indices pertaining to these instruments—conducts a formal auction of the senior, unsecured debt of defaulted issuers.

The corporate indices trade with a fixed spread (coupon), which is calculated just prior to a roll date by averaging the quotes from participating dealers. The spread, the premium that protection buyers pay to protection sellers, represents the coupon that generates a zero net present value on the contract's expected cash

Exhibit 13.4 Actively Traded CDS Indices

	CDX.NA.IG	**iTraxx Europe**	**ABX.HE**	**CMBX.NA**	**LCDX.NA**
Type of asset	Senior, unsecured debt	Senior, unsecured debt	Residential MBS bonds	Commercial MBS	Secured, First lien loans
Location	North America	Europe	U.S.	North America	North America
Number of entities	125	125	20	25	100
Settlement	Physical	PAUG	PAUG	PAUG	Physical/Cash
Ratings	Investment grade	Investment grade	Investment grade	AAA—BB	BB and B
Inception	October 2003	June 2004	January 2006	March 2006	May 2007
Quote type	Spread	Dollar pPrice	Dollar price	Spread	Dollar price
Subindices	Industrial Financial High-volatility Telecom	High-volatility crossover Financials Nonfinancials Industrials Energy; Auto Consumer; TMT	Aaa Aa2 A2 Baa2 Baa3	AAA AA A BBB BBB– BB	
Tranche attachments	3%, 7%, 10%, 15% 30%	3%, 6%, 9%, 12% 22%			

flows at issuance. As the market reassesses individual single-name CDS risks, the spread may change. If a CDS trade is initiated after an index is first priced, then an up-front payment or discount is applied. The protection buyer pays premiums every quarter on fixed payment dates and a prorated partial payment if the contract is initiated between payment dates. The protection buyer also makes an accrued premium payment if one of the underlying assets defaults on a nonpayment date.

The corporate indices use standardized maturities (5, 7, and 10 years are most liquid) and regular payment dates, usually the twentieth of March, June, September, and December each year. Off-the-run index series continue to trade but with reduced liquidity. Under normal market conditions, the corporate indices and subindices are very liquid, with tight bid-ask spreads, averaging less than 1/2 of a basis point. These bid-ask spreads widened considerably when the subprime crises infected corporate credit markets in 2007 and 2008.

An alternative to the investment-grade corporate indices described above, LCDX references a basket of 100 syndicated, first lien, secured, leveraged loans. Launched in June 2007, LCDX promised to expand the market for collateralized loan obligations (CLOs) by offering issuers an opportunity to source assets synthetically. Unfortunately, market turmoil and reduced demand for structured products prevented LCDX from fulfilling its promise. Still, some investors welcomed the opportunity to hedge CLO positions with the index. Like its U.S. investment grade cousins, LCDX defines two credit events only, bankruptcy and failure-to-pay. The index has a five-year term and uses cash settlement protocol.

The most actively traded structured product indices include ABX.HE and CMBX. ABX.HE references a basket of 20 asset-backed security issues collateralized by subprime HEL. CMBX references a basket of 25 commercial mortgage-backed security issues.

ABX.HE comprises five subindices created by pooling like-rated tranches with ratings of AAA, AA, A, BBB, and BBB–. The first series, 06–01, was launched in January 2006, and new series are introduced every six months, in January and July. However, the January 2008 roll was canceled due to the collapse of the subprime loan market and the dearth of new issuance.

ABX.HE indices trade with the standardized pay-as-you-go template, which allows for three floating amount events: interest shortfalls (fixed cap only), principal shortfalls, and write-downs (or implied write-downs). Write-downs and principal shortfalls trigger a payment from the protection seller to the protection buyer. The index notional amount adjusts as any of the reference entities amortize, prepay, default, or incur write-downs or write-ups.

Market quotes for ABX.HE are price based rather than spread based. When a series rolls, its initial coupon (fixed premium) and price are determined. Because the initial coupon is capped at 5 percent, the initial price may be set below par. As the market reassesses individual single-name CDS risks, the price may deviate from its initial value. If an index trade is initiated when the quoted price is below par, the protection buyer must make an up-front payment to the protection seller equal to par minus the index price. Alternatively, if a trade is initiated when the quoted price exceeds par, the protection seller must make an up-front payment to the protection buyer equal to the index price minus par. Protection buyers make regular monthly payments to protection sellers equal to one-twelfth of the product of the current notional amount of the index and the fixed premium established on the roll date.

Dealers derive significant benefits from indices that are actively traded. New CDS indices are routinely introduced; some succeed while others fail. The most successful indices are those that are generic enough to appeal to a large investment audience yet specific enough to allow for good two-way order flow between hedgers and speculators. In general, indices that are not representative of the cash market fail as hedging instruments.

TRANCHES OF CDS INDICES

Some CDS indices (e.g., CDX, iTraxx, and ABX) are sliced into tranches, replicating a synthetic CDO with the index's reference entities acting as collateral. Index tranches allow broker-dealers to offset the risks they incur when they structure transactions for investors who target a particular credit strategy to a specific point in the capital structure. A tranche is defined by attachment and detachment points. The attachment point defines the amount of subordination a tranche enjoys. The tranche thickness, measured by subtracting the attachment point from the detachment point, represents the maximum loss that can be sustained. For example, consider a 7 to 10 percent tranche. This tranche first experiences losses when the index suffers losses in excess of 7 percent of the notional. The tranche can then withstand losses of an additional 3 percent (the tranche thickness). As with any standard CDS index, a tranche can be traded after an initial transaction by unwinding the contract or by assigning it to a new counterparty.

Credit events and tranche characteristics such as maturity, payment dates, and premium accruals, follow the conventions of the CDS index in question. In general, protection buyers pay sellers either a full running spread or a combination of a running spread and an up-front payment. The latter convention is commonly used for tranches that are more likely to incur losses, such as the first loss, 0 to 3 percent tranche of an entity the market believes is at risk of bankruptcy. Protection sellers compensate buyers for any losses in the index portfolio that breach the tranche's attachment point. Losses that exceed this attachment point reduce the notional amount of the tranche on which the spread is paid.

As an example, consider the 7 to 10 percent tranche of CDX.IG. The attachment point is 7 percent the detachment point is 10 percent, and the tranche width is 3 percent. Assume the index notional is $1 billion. The tranche would have an initial notional of $30 million. It would start to sustain losses when portfolio losses exceed $70 million (7 percent of $1 billion). The tranche's notional would be reduced to zero at $100 million losses (10 percent of $1 billion). It is obvious that tranches with lower attachment points are more likely to sustain losses. To compensate for this risk, spreads tend to be higher on tranches that are lower in the capital structure. For two tranches with the same attachment point but different detachment points, the thinner tranche (i.e., the one with the lower detachment point) faces a higher probability of a complete write-down and is therefore more levered (see Exhibit 13.5).

The distinction between losses and defaults is critical and illustrative of the difference between the ratings of Standard & Poor's and Moody's. The former rates to the first dollar loss while the latter rates to expected loss. Tranche attachment

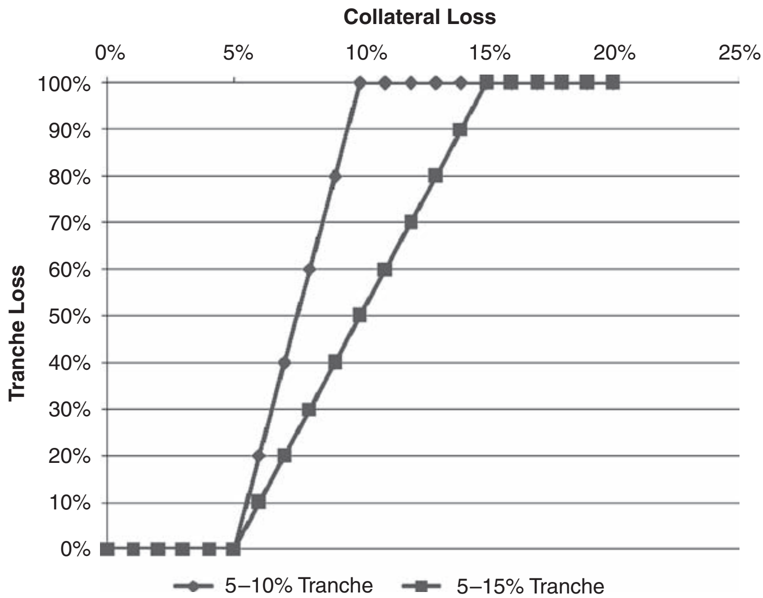


Exhibit 13.5 Tranche Loss Given Collateral Loss for 5 to 10 Percent and 5 to 15 Percent Tranches

and detachment points refer to portfolio losses, not defaults. Assuming a loss given default of 50 percent, the 7 to 10 percent tranche can withstand defaults of up to 14 percent of the portfolio ($14\% \times 50\% = 7\%$, the attachment point) before it sustains losses. Since the underlying CDX index aggregates 125 corporate entities, the tranche would experience no cash flow losses for the first 17 defaults, though its credit support would shrink from 7 percent to 0.2 percent ($7\% - 50\,\% \times 17/125$). The eighteenth default would cause the tranche to lose 6.667 percent of its notional [$((50\% \times 18/125) - 7\%) \,/\, 3\%$]. If 26 defaults occur, the tranche's notional would be entirely wiped out, and the note above the 7 to 10 percent tranche would incur its first dollar loss.

These results are based on a static recovery rate of 50 percent. In reality, recovery rates tend to be highest during periods of low default and lowest when defaults reach historical peaks.[3] The results also offer little information about returns or price volatility. Additional information on the timing of defaults is needed to compute these metrics. Clearly, if all the defaults occur near the end of a contract's life, the protection seller would fare better because she would receive premiums for a longer period of time. For this reason, investors need to consider not only loss expectations, but also the timing of defaults. *Correlation* is the term used to describe the degree to which defaults are synchronized among collateral names, and tranche trading strategies are often called correlation trades. We address correlation issues in the final section of this chapter.

Synthetic ABS and CMBS indices were launched in early 2006. Standardized tranches for these indices appeared soon after. Thus far, these indices and tranches have not garnered the same success as their corporate cousins, at least when liquidity is our metric for performance. Although the ABX had a fighting chance for liquidity before the subprime crisis of 2007, there are two reasons why we should expect CDS on structured products to be less liquid, even in the best of times.

1. The underlying cash markets are smaller and less liquid than for corporate bonds.
2. The PAUG template necessitates a cash flow model for structured products.

Vendor software is available, but cash flows are extremely sensitive to model assumptions about prepayments. The dealer community was enthusiastic about TABX, a tranched version of ABX, launched in February 2007. However, TABX has not lived up to expectations, due to its limited diversification and lack of liquidity.

TRADING STRATEGIES USING INDEXES AND TRANCHES

Synthetic assets such as CDS index tranches, CDS indices, and single-name CDS (on corporate debt, ABS and CDOs) theoretically add to market completeness, allowing investors to speculate on or hedge against fairly nuanced outcomes. Using credit indexes and tranches, investors can efficiently hedge assets in inventory against spread movements or defaults. CDO issuers can populate a new security or hedge the ramp-up risk of a new issue. Investors can express macro views and make relative value (convergence) trades across asset classes, vintages, geographical regions and the capital structure. We provide a few examples next.

Hedge assets against spread movements or default: An investor who is long an AA-rated tranche of a high-grade ABS CDO could buy protection on that tranche in the single-name CDO CDS market.

Ramp a CDO: A CLO manager could quickly gain exposure to a basket of leveraged loans by selling protection on LCDX.

Express a macro view: An investor who is bullish on leveraged loan performance could sell protection on LCDX; an investor who is bearish on commercial mortgage-backed security valuations could buy protection on CMBX.

Relative Value Trades

Across sectors: An investor who believes asset-backed securities populated by home equity loans are expensive relative to the investment-grade bonds of American financial companies could sell protection on the financial sector of CDX.NA.IG while simultaneously buying protection on the ABX.

Across vintages: Since the structured product indices generally reference specific assets that were originated within 6 months of the roll date, it is possible to pit the performance of one vintage against another. An investor who believes subprime assets that were originated in the first half of 2006 are superior to those originated in the second half of 2007 could sell protection on ABX 06–01 and buy protection on ABX 07–02.

Across geographical sectors: An investor who believes European growth will be weaker than U.S. growth could sell protection on CDX.NA.IG and buy protection on iTraxx Europe.

Across the capital structure (i.e., a correlation trade): An investor who believes systematic risks are on the rise could initiate a correlation trade by selling protection on the 0 to 3 percent tranche of CDX and buying protection on the 3 to 7 percent tranche of CDX. Such a trade could be made indifferent to small changes in the reference spread by sizing the relative positions so that the combined trade has a delta equal to zero. Here, delta measures a tranche's profit and loss sensitivity to small changes in the average spread of the underlying portfolio of assets. We discuss delta hedging in the section on correlation.

MARKET DYNAMICS: CDS AND CDOS

New-issue CDO spreads are influenced by the demands of long-term investors and the supply of new securities. These in turn are affected by the activities of at least three distinct groups: speculators, hedgers, and arbitrageurs. Speculators (often hedge funds) may choose to express a bullish or bearish view on an asset class by selling or buying protection on a CDS index. Dealers who are naturally long securities due to their market-making activities, often hedge their inventory or new-issue pipelines by taking short positions in the tranches of CDS indexes. Additionally, certain long-term investors may seek to hedge their CDO holdings with single-name CDS or indexes. Meanwhile, arbitrageurs (often dealers and CDO managers) will attempt to exploit the price differences between the cash and synthetic markets. However, the "arbitrage" opportunities in structured products markets are never without risk. Markets can freeze, bid-ask spreads can widen and basis risk cannot be completely eliminated.

It is common to fixate on one dominant group that seems to be moving the market. In reality, every trade involves two parties, often with opposing views. In general, when bearish sentiment is strong, CDS spreads trade wide to cash and synthetic bid-ask spreads widen. When bullish sentiment dominates, CDS spreads trade tight to cash and the bid-ask spread narrows. CDO cash spreads are stickier than CDS spreads or index prices because it is not possible to take a short position in the cash market. Hence, CDO cash spreads tend to lag the synthetic markets.

SYNTHETIC CDOS AND BESPOKES

Synthetic CDOs combine CDO and credit default swap technologies. The first synthetic CDOs were created in 1997 and referenced specific assets on bank balance sheets. These structures allowed banks to free up capital by selling to investors the risk of default on a basket of assets. Many of the original structures were blind pools, which meant that investors did not know the specific credits but instead were informed about the aggregate collateral statistics. These structures quickly went out of favor, as banks often bundled their worst assets and investors suffered poor performance.

In a synthetic CDO, a special-purpose vehicle (SPV) serves as an intermediary between buyers and sellers of credit risk. The SPV acquires risk via CDS contracts instead of cash assets. The SPV sells protection on a reference basket of assets to a sponsor (or arranger). It receives a premium for the risk it assumes. The buyer of protection pays that premium. The SPV can distribute the credit risk to different tranches. The tranches will receive a portion of the total premium, based on the net risks assumed by the tranches.

A transaction is fully funded when the sponsor raises capital from investors and purchases high-quality assets (see Exhibit 13.6). Eligible high-quality assets include triple-A rated money market debt, government or agency notes, and guaranteed investment contracts (GICs) issued by highly rated insurance companies. In a fully funded transaction, investors may be exposed to dealer counterparty risk if the collateral securitizing a transaction is held by the dealer or monoline counterparty risk if funds are invested in a GIC issued by a monoline.

The default swap premiums paid by the protection buyer, combined with the cash flows from the high-quality assets are available to be distributed to the note investors, either on a prorated basis or sequentially, from the top of the capital structure to the bottom. Note investors are essentially entering into a swap agreement that pays them LIBOR plus a fixed premium on a notional amount. That notional amount could be reduced to zero if the losses on the reference assets exceed the note's detachment point.

Equity investors are essentially selling protection on the collateral pool and simultaneously buying protection above a particular attachment point from the note investors. When any asset in the collateral pool experiences a credit event, the SPV pays the protection buyer an amount linked to the loss incurred on the stressed asset. The loss is then passed on to investors in reverse order of seniority, with the equity tranche absorbing the first loss.

A tranche's risk profile is related to its attachment and detachment points. As with index tranches such as CDX, the attachment point defines the cushion of subordination a tranche enjoys before it is vulnerable to its first dollar loss. With each dollar loss above its attachment point, a tranche experiences a dollar decline in its notional amount. The detachment point defines the total collateral losses that would have to occur for a tranche to experience a complete loss of principal. A tranche's thickness, the difference between its detachment and attachment points, measures the total amount of collateral losses the tranche could sustain. As

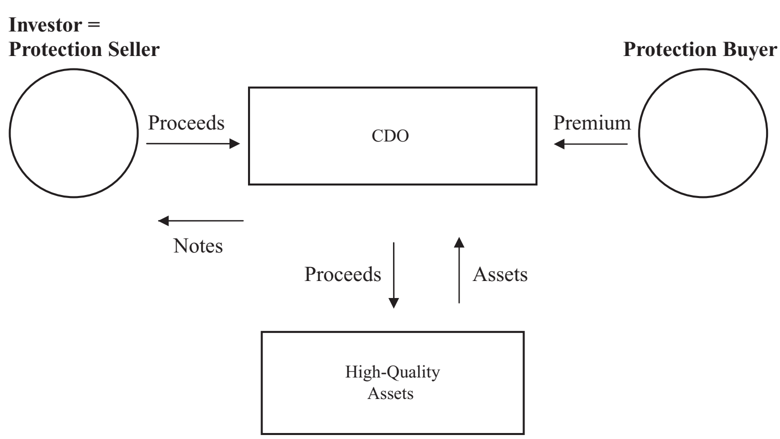


Exhibit 13.6 Fully Funded Synthetic CDO

previously stated, given two tranches with equivalent attachment points, the thinner tranche is more levered and riskier.

Bespokes are synthetic CDOs, formed when an investor, working with an arranger (dealer), selects a portfolio of reference credits and specifies a pair of attachment/detachment points for a customized tranche. Bespokes are simpler than traditional CDOs, which rely on a cash flow waterfall mechanism and various quality and coverage tests to determine how to divert cash interest and principal payments when the collateral becomes stressed. The development of liquid index tranches has facilitated the growth of bespoke issuance, allowing dealers to partly hedge the residual risks they often retain when they structure synthetic CDOs.

Synthetic CDOs enjoy three advantages relative to cash CDOs.

1. Structures tend to be clean and simple, with no complex waterfalls. Synthetic CDOs usually dispense with interest coverage and overcollateralization tests.
2. Synthetic CDOs enjoy efficiencies in ramp-up, asset selection, and funding costs, especially at the top of the capital structure (e.g., the super-senior tranche). The reference portfolio can be set up instantly and is not limited to assets that are available for cash trading. Moreover, investors can take short positions in certain assets, something that is difficult to do in the cash market.
3. Synthetic assets eliminate the prepayment and interest rate risks of cash assets, simplifying investors' efforts to hedge currency risks with swaps.

Compared to cash CDOs, synthetic CDOs have three disadvantages.

1. Most investors must mark-to-market their positions because synthetic tranches are classified as derivatives. A brief or extended period of high market volatility could result in dramatic mark-to-market losses.
2. Synthetic CDOs may not enjoy the same level of liquidity as cash CDOs, due to reduced information flow. Cash CDOs are typically modeled by data providers such as Intex, and an offering memorandum is usually available. Synthetic CDOs tend to be private transactions executed by one originator for one investor. The originator usually will make a market in the security, but there is no guarantee other dealers will do the same.
3. Synthetic CDOs may be penalized by the more liberal credit event definitions present in the standardized templates for credit default swaps on corporate debt, ABS, and CDOs. Compared to cash assets, losses may be front-loaded due to implied write-downs, distressed rating triggers, and certain PIK events (see Exhibit 13.7).

CORRELATION

Correlation is a term used to describe the degree to which asset defaults are synchronized and correlation trading strategies involve the buying and selling of index or synthetic tranches. The best way to illustrate correlation is with an example.

Exhibit 13.7 Synthetic CDOs and Bespokes: Advantages and Disadvantages

Advantages	Disadvantages
Clean structures	Positions must be marked to market
Instantaneous ramp-up	Reduced liquidity
Expanded universe of assets	PAUG template may front-load losses
Cheaper super-senior funding costs	
Simplified currency risk hedging	

Consider a portfolio containing 50 investment-grade corporate bonds with an average yield spread of 130 basis points over LIBOR. Suppose we carve the portfolio into three slices, equity, mezzanine, and senior notes, with attachment—detachments points of 0 to 15 percent, 15 to 40 percent, and 40 to 100 percent respectively. In Exhibit 13.8, we see that the average spread of 130 basis points is shared unequally among the tranches. Equity receives the largest share (75 basis points), and the mezzanine and senior notes receive much smaller shares (25 and 30 basis points respectively).[4]

The average spread tells us something about how many defaults we should expect within the portfolio. The tranche spreads tell us something about the likelihood that bond defaults will be synchronized in time. In this example, equity receives the lion's share of the spread, so we can conclude that defaults are not likely to be highly coordinated. Otherwise, the mezzanine and senior tranches would require higher spreads to compensate investors for the increased probability that losses would hit these tranches.

Correlation conveys how risk is distributed across the tranches (capital structure). Portfolios with low correlation have risk concentrated in the first-loss tranche, whereas portfolios with high correlations allocate more risk (and spread) to the mezzanine and senior notes. Correlation measures the degree to which cross-sectional defaults are synchronized. More precisely, correlation measures the relationship between *pairs* of random variables, such as bond default times. Most CDO and synthetic pricing models assume a common correlation value for all pairs of assets within a sector. Hence we speak of one correlation metric for a portfolio of like assets.

With high correlations, defaults are coordinated; conditional on one default, other defaults are likely. With low correlations, defaults diverge. Note that correlation says nothing about default probabilities; it is entirely possible to have low default rates and high correlations.

Exhibit 13.8 Tranche Spreads

Attach—Detach Point	0 to 15%	15 to 40%	40 to 100%	Total**
Spread	500	100	50	
Spread x Thickness	75	25	30	130

**Total spread equals the sum of the products of the tranche spreads and the tranche thicknesses.

Assume that each of 50 corporate bonds in our portfolio has a 10 percent probability of defaulting over the next year. Now suppose the bonds are virtually identical. Then they are perfectly correlated, and there are only two possible outcomes: Either all 50 bonds default, or none defaults. With perfect correlation, the probability of 50 defaults is 10 percent and the probability of 0 defaults is 90 percent. Also note that although the expected number of defaults is 5 [$(50 \times .10) + (0 \times .90)$], the outcome is highly variable.

Now suppose each bond is unique and default behavior is independent. With zero correlation, the probability of 50 defaults is $.10^{50}$ and the probability of 0 defaults is $.90^{50}$, both low-probability events. Most of the probability distribution is in fact centered near the mean number of defaults, which is still five.[5] With 0 percent correlation, the outcome has low dispersion, or low volatility (see Exhibit 13.9).

Default correlations are usually positive for most pairs of assets because systemic risks, such as the risk of a recession, inflationary spike, or liquidity crisis impact assets similarly (in the same direction). Continuing with our example, suppose investors reassess economic risks and conclude that default correlations will decline going forward. If we assume total risk does not change, then the portfolio average spread remains unchanged, but there is a repricing of risk across the tranches. With lower correlations, there is a reduced chance that the senior notes will incur a loss. There is also an increased chance that equity will suffer a poor outcome.[6] Thus, in a low correlation environment, spreads on equity could rise to 600 and spreads on senior notes could fall to 25.

Holding spreads (or average risk) constant, a decline in correlation raises the value of a long senior-note position (i.e., spreads fall) and lowers the value of a long

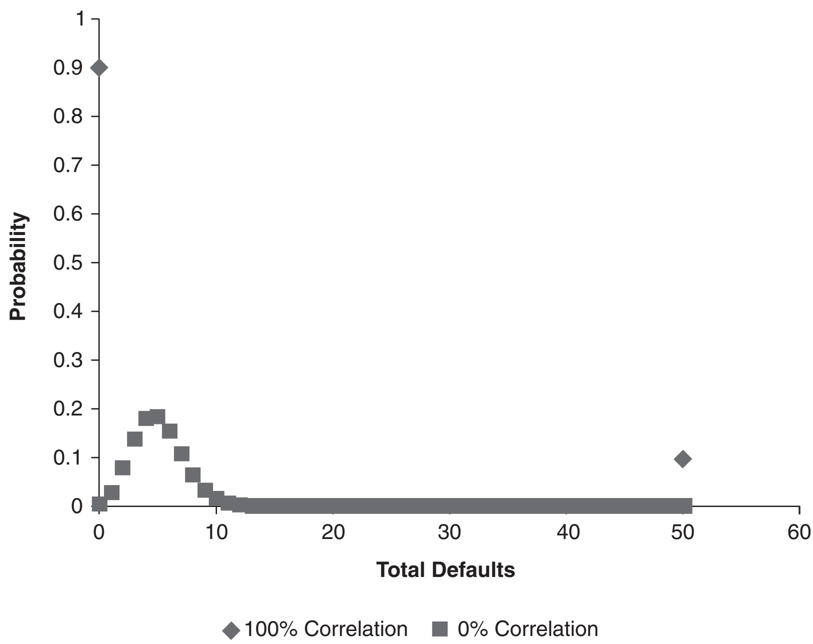


Exhibit 13.9 Probability Distribution for Total Defaults

equity position (i.e., spreads rise). In our example, the mezzanine tranche faces no price change, but this outcome is by design.

Now assume we move in the opposite direction and investors conclude that default correlations will increase going forward. If total risk remains unchanged, then risk is reallocated across the capital structure, with spreads on equity falling to 300 and spreads on senior notes rising to 100. Holding spreads (or average risk) constant, an increase in correlation raises the value of a long equity position (i.e., spreads fall) and lowers the value of a long senior note position (i.e., spreads rise)[7] (see Exhibit 13.10).

Default correlations vary through time, rising and falling with the economic cycle. In general, during good times, idiosyncratic risks outweigh systemic risks, and correlations are low. During bad times, systemic risks dominate, and correlations tend to rise. Dramatic changes in correlation values do not usually occur without coincident large moves in spreads. This is because the macroeconomic factors that affect correlation also influence credit spreads. However, due to technical factors related to bespoke issuance and dealer hedging activities, CDOs may occasionally experience a significant repricing of correlation. On the whole, senior tranches provide exposure to market-wide systemic risks while equity tranches provide investors exposure to idiosyncratic risks.

In our simple example, the mezzanine tranche is correlation independent. This is generally true for small correlation moves. For larger shifts in correlation, the mezzanine tranche behaves differently. When spreads tighten dramatically, the mezzanine tranche resembles a senior note; when spreads widen dramatically, the mezzanine tranche performs like equity.

Recall that delta measures a tranche's profit and loss sensitivity to small changes in the average spread. Gamma (or convexity) risk measures the change in delta per unit change in the underlying asset spread. Generally, gammas are positive for equity tranches and negative for senior tranches (see Exhibit 13.11). From the options world, we know that when an option is deep in-the-money or deep out-of-the-money, its convexity is approximately zero.

When we combine two tranches (or one tranche and the index) in such a way as to neutralize the delta, the gamma of the combined position will almost always be nonzero. For example, if we take a long position in an equity tranche and a short position in the index, the combined position will be positively convex, benefiting from increased volatility. As spreads widen, delta for the equity tranche declines. If we do not rebalance the portfolio, there will be a loss on the long equity position and a gain on the short index position. The gain on the short index position will

Exhibit 13.10 Tranche Spreads under Different Correlation Assumptions

Attach—Detach Point	**0 to 15%**	**15 to 40%**	**40 to 100%**	**Total****
Base Case Spread	500	100	50	130
Low-Correlation Spread	600	100	25	130
High-Correlation Spread	300	100	100	130

**Total spread equals the sum of the products of the tranche spreads and the tranche thicknesses.

Exhibit 13.11 Tranche Delta and Gamma Statistics

Tranche	Attach—Detach	Delta	Gamma
Equity	0 to 10%	4.0	0.30
Mezzanine	10 to 30%	1.0	0.06
Senior	30 to 100%	0.57	–0.09
Index	0 to 100%	1.00	0.00

**Total spread equals the sum of the products of the tranche spreads and the tranche thicknesses.

exceed the loss. In this situation, we are effectively overhedged. Conversely, if spreads tighten, we will experience a gain on the long equity position and a loss on the short index position, with the gain exceeding the loss. Here we are effectively underhedged.

Trading desks can be long senior tranches as a natural consequence of selling mezzanine and equity bespokes. Suppose a desk has sold protection on a senior tranche (i.e., it is long senior tranche risk). Assume the desk delta-hedges its position using the index. If spreads widen (systematic risk rises), the desk is underhedged. The loss on the tranche will exceed the gain on the index. Spread widening will force the desk to buy more protection to maintain an effective hedge. If many trading desks hold similar positions, spreads may widen considerably, creating increased hedging activity, wider spreads, increased protection buying, and so on. This vicious cycle can feed on itself and lead to a dramatic repricing of risk and correlation.

The industry standard models that are used to price the tranches of credit baskets assume all reference credits share the same pair-wise default time correlation coefficients. Defaults arrive via a Poisson process with static intensity; postdefault recoveries are non–time varying and identical for all assets. Using the observed market prices for credit baskets and tranches, the industry standard model can be used to back out an implied correlation coefficient, in much the same way one uses a Black-Scholes model to compute an implied volatility. There is one very important difference, however. For equity options, volatility is directly observable and hedgeable. In contrast, default correlation is neither observable nor hedgeable in single-tranche CDOs. Moreover, correlation risk can be quite significant.

In early 2008, the industry standard models were on the verge of failing, with some credit basket prices implying a correlation coefficient above 1. Market participants tried to address this problem by lowering the assumed recovery rate. This worked temporarily, but the flat capital structure for correlation still presented problems for risk managers.

There is a large literature devoted to the question of how to value credit derivatives, such as index tranches. Two general pricing paradigms prevail: structural models and reduced form (or intensity) models. Structural models assume that defaults occur endogenously, when the value of a firm's assets falls below the firm's outstanding debt. This approach is taken by Merton (1974), Black and Cox (1976), and Longstaff and Schwartz (1995), among others. Typically, structural models assume firm value follows a stochastic process. Asset recoveries, postdefault, are typically a simple function of structural variables. To model and price credit baskets, asset correlation values based on historical data can be used. Structural models are economically motivated but often present estimation challenges, especially for firms that do not have publicly traded equity. Calibrating the term structure of credit default swap spreads can be daunting. In general, structural models are best suited for investment and rating decisions.

In contrast to structural models, reduced form models assume defaults occur exogenously. This is the approach taken by Jarrow and Turnbull (1995), Li (1999), Duffie and Singleton (1999), and Longstaff and Rajan (2008), among others. It is also the industry standard approach. With reduced form models, state variables can be introduced to create dependencies between interest rates, default intensities, and recovery rates. Although the simplest models do a poor job pricing traded credit baskets, models that incorporate jump processes, such as that of Davis and Lo (2001), or catastrophic events, such as that of Longstaff and Rajan (2008), show more promise. Compared to structural models, reduced form models seem ad hoc, as there is no economic rationale for defaults. However, reduced form models can be more tractable than their structural counterparts. In general, reduced form models are best suited for hedging and risk management.

CONCLUSION

Credit derivatives are a growing presence in today's capital markets, fueled in large part by increased cross-border flows of capital and deregulation. As capital markets become more integrated and developing economies accumulate greater wealth, credit derivatives will continue to evolve, serving the needs of an expanding investor base. Though short-term liquidity concerns and counterparty failures may lead to temporary market disruptions, credit derivatives will survive.

Credit derivatives fulfill the basic needs of hedgers who seek risk-reduction strategies and investors who seek nuanced investment strategies. Going forward, we can expect greater innovation and a host of new products. Many will fail, but some will find a large audience and succeed. Investors should strive to understand credit derivatives because these synthetic assets increasingly influence the trading behavior and prices of cash assets.

ENDNOTES

1. Leveraged loans are bank loans extended to noninvestment-grade borrowers. The loans are collateralized by real assets and generally enjoy higher recoveries than unsecured corporate bonds. Leveraged loans can be broadly syndicated (issued to medium-size companies and traded by a consortium of large commercial and investment banks), or so-called middle market loans made to smaller companies (those with earnings before interest, taxes, depreciation, and amortization [EBITDA] of $50 million or less).
2. See Chapter 14 for more detail. A sample waterfall appears in Exhibit 13.8 and a sample O/C test appears in Exhibit 13.10.
3. See Deconstructing CDOs, Wachovia Capital Markets, LLC, Brian McManus, Dave Preston, Steven Todd, and Anik Ray, May 23, 2007.
4. Theoretically, there is an arbitrage when the index spread does not equal the sum of the tranche weighted average spreads (where the weights correspond to tranche thickness

levels). However, wide bid-ask spreads may make it difficult to take advantage of this arbitrage.

5. The total number of defaults follows a binomial distribution, with p, the probability of default equal to 10% and n, the number of trials, equal to 50.
6. In this example, with 100% correlation, there is a 90% probability that equity suffers no loss (See Exhibit 13.9). In contrast, with 0% correlation, there is more than a 99% probability that equity suffers some loss.
7. A distinction is made between tranche (compound) correlation and base correlation. With base correlation pricing, correlations are computed for contiguous segments, from the bottom tranche up. In our example, we could compute correlations for the following three segments: 0 to 15%, 0 to 40% and 0 to 100%. In general, the curve of correlations obtained by calibrating to the first loss tranche is more stable than that obtained by computing correlations for detached segments.

REFERENCES

The Bank for International Settlements, www.bis.org

Black, Fischer and John Cox. 1976. "Valuing corporate securities: some effects of bond indenture provisions," *Journal of Finance* 31: 351–367.

Chaplin, Geoff. 2005. *Credit Derivatives: Risk Management, Trading and Investing* Hoboken, NJ: John Wiley & Sons, Ltd.

Davis, Mark H.A., and Violet Lo. 2001. "Infectious defaults," *Quantitative Finance* 1: 382–386.

Duffie, Darrell and Kenneth Singleton. 1999. "Modeling term structures of defaultable bonds," *Review of Financial Studies* 12: 687–720.

"Integrating Credit Derivatives," Brian McManus, Steven Todd, Anik Ray, David Preston and Justin Pauley, Wachovia Capital Markets, LLC, October 4, 2007.

The International Swaps and Derivatives Association, Inc., www.isda.org

Jarrow, Robert and Stuart Turnbull. 1995. "A Markov model for the term structure of credit risk spreads," *Review of Financial Studies* 10: 481–523.

Li, David. 1999. "On default correlation: a copula function approach," Working Paper 99–07, The Risk Metrics Group.

Longstaff, Francis and Arvind Rajan. 2008. "An empirical analysis of the pricing of collateralized debt obligations," *Journal of Finance* 63: 529—563.

Longstaff, Francis and Eduardo Schwartz. 1995. "A simple approach to valuing risky and floating rate debt," *Journal of Finance* 50: 781–819.

Markit, www.markit.com

McManus, Brian *et al.* 2007. "Deconstructing CDOs," Wachovia Capital Markets, LLC, May 23, 2007.

Merton, Robert. 1974. "On the pricing of corporate debt: the risk structure of interest rates," *Journal of Finance* 29: 449–470.

The Securities Industry and Financial Markets Association, www.sifma.org.

ABOUT THE AUTHOR

Steven Todd is an associate professor of finance at Loyola University Chicago. Dr. Todd completed his Ph.D. in finance and business economics at the University of Washington; he also holds a B.S. degree in operations research from Cornell University. Dr. Todd teaches graduate and undergraduate courses in corporate finance, investments, options, derivatives, credit risk management, international finance, portfolio management and investment banking. His research interests include asset pricing, financial markets and institutions and corporate governance. He has published papers on securitization, mutual fund performance measurement, managerial compensation, and equity analysts in several journals, including *Real Estate Economics*, *Journal of Business*, *Journal of Corporate Finance*, and *Financial Markets*, *Institutions and Instruments*. Prior to his academic career, Dr. Todd worked on Wall Street as an analyst, investment banker, and trader at Merrill Lynch and UBS Securities. Recently, Dr. Todd served as a director in the CDO Research Group at Wachovia Securities.

CHAPTER 14

Structured Credit Products

STEVEN TODD
Associate Professor of Finance, Loyola University, Chicago

Structured products include notes linked to interest rates, equities, foreign exchange and commodity assets, as well as credit derivatives, such as asset-backed securities (ABS), collateralized debt obligations (CDOs), and commercial mortgage-backed securities (CMBS). This chapter focuses on the latter group, known as structured credit products. These securities are created by pooling various assets and allocating the cash flows from the pool of assets to various tranches, classes, notes, bonds, or other securities. The assets that are pooled can include virtually any traded or nontraded assets, including consumer loans (residential mortgages or credit card, auto, and student loans), corporate debt and equity securities (leveraged loans, commercial mortgages, corporate bonds, common and preferred equity), government debt (e.g., emerging market sovereign debt), commodities, or other structured products.

One feature common to all structured credit products is the use of financial engineering techniques to create securities that provide a range of risk-return profiles for different investors. Credit enhancement in the form of subordination or overcollateralization can transform high-yield or unrated assets into a family of securities with varying credit risks, from triple-A-rated notes to unrated, highly levered, equity-like securities.[1]

Structured credit products differ from other structured securities that partition interest rate risks instead of credit risks. Included in this asset class are mortgage-backed securities (MBS) and collateralized mortgage obligations (CMOs), securities that reference residential mortgage loans, usually, prime, fixed rate conventional conforming loans or government-guaranteed loans.[2] When these securities are issued by the federal housing agencies (Government National Mortgage Association, Federal National Mortgage Association, and Federal Home Loan Mortgage Corporation), they enjoy implicit government guarantees that render them virtually free of credit risk.[3] When these securities are issued by private entities, such as commercial and investment banks, they usually incorporate some form of private insurance to eliminate credit risk. Currently, MBIA and Ambac are the largest underwriters of bond insurance.

Exhibits 14.1 and 14.2 provide summary data on the amount of securities outstanding for various categories of structured products. Note that agency MBS and CMOs dominate the charts. Part of the reason for this is that MBS and CMOs were the original prototypes for structured products. These securities were created

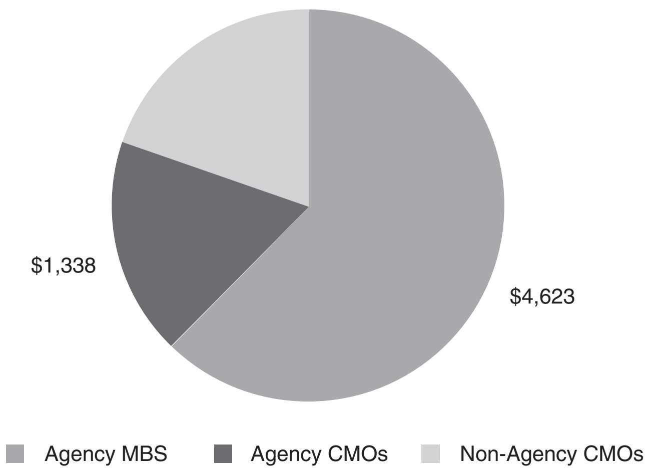


Exhibit 14.1 Mortgage-Related Securities Outstanding $ Billions (as of 2008: Q3)
Source: Securities Industry and Financial Markets Association.

in the late 1970s and early 1980s. Structured credit products followed in the 1990s. Also note that asset-backed securities (which include several sectors) are the most prolific structured credit product.

Structured credit products usually are designed as trusts or special-purpose vehicles (SPVs) that are remote and nontaxable, providing certain legal advantages to the issuer, such as limited liability and immunity from lawsuits. During the market turmoil of 2007, some of these features were tested as subprime assets experienced more defaults and losses than the rating agencies considered possible.

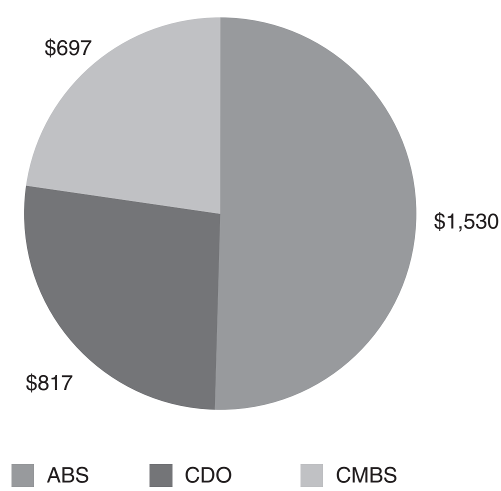


Exhibit 14.2 Structured Credit Products Outstanding $ Billions (as of 2006: Q4)
Source: Securities Industry and Financial Markets Association.

Structured credit products offer term funding (financial capital at a fixed rate over a fixed period) to issuers, a benefit that has contributed to their appeal and growth. Banks and other financial institutions often issue securities to better manage their balance sheets and diversify their sources of funding, which typically include deposits, loans, bonds, and equity. Structured credit products appeal to asset managers because they allow managers to increase their assets under management, generating additional fee income.

In the primary market, structured products often are created by large commercial banks, investment banks, or financial firms that have access to assets that can be securitized. Sometimes underwriters will structure what are called arbitrage transactions. Here the net value of the securities (liabilities) that can be created and sold exceeds the total cost of the assets that collateralize an issue. Arbitrage is a poor name, however, because there is no guarantee that a profit will be made, as market prices can change quickly before a transaction is priced and sold.

Structured credit product investors include financial institutions (commercial banks, insurance companies, pension funds, and money managers), hedge funds, and sovereign funds around the globe. Some structured credit products are privately issued securities, available for purchase by accredited, qualified institutional investors only. Others are publicly traded. Much of the issuance and investment activity is centered in the United States, but increasingly Europe, Asia, and Latin America are becoming involved. Secondary market trading is quite developed for some sectors but sporadic for others.

ABS, CDO, and CMBS assets would not have grown so quickly if they did not offer investors real advantages. Among the benefits these assets bring to investors is regulatory arbitrage, namely, the opportunity to invest in investment-grade securities that reference unrated assets that otherwise would not be eligible investments. Many institutional investors are constrained from investing in assets below a minimum credit rating. By giving investors access to an expanded set of assets, structured credit products offer diversification benefits. Additionally, they allow investors to custom-tailor risk-return strategies. For example, CDOs usually are structured as floating rate securities. As such, they allow investors to separate credit risk from interest rate risk. Moreover, the cash flows from CDOs are easily swapped into foreign currencies, allowing foreign investors the opportunity to hedge exchange rate risk.

Although structured credit products are mostly credit portfolios offering investors exposure to credit risk in various forms, there are often other embedded risks, which can be hedged or not. These include interest rate risk in the form of caps and floors, prepayment risk (especially for negatively convex securities such as residential mortgage loans),[4] early termination risk (as when optional redemptions or auction/cleanup calls are exercised), foreign exchange risk, counterparty risk (especially in securities that reference assets synthetically via credit default swaps), and basis risk (e.g., when an asset is indexed to one benchmark rate and a liability references another, similar rate). In general, because interest rate and credit cycles are asynchronous, interest rate risks can be used to partly hedge credit risks.

In the next three sections, we examine the three most actively issued structured credit products: ABS, CDOs, and CMBS.

ASSET-BACKED SECURITIES

Asset-backed securities reference consumer assets, such as mortgage loans, credit card receivables, auto loans or leases, and student loans. The largest sector of the ABS market is the residential mortgage sector, known as the home equity loan (HEL) market, because the earliest transactions were collateralized by second lien mortgage loans to prime borrowers (see Exhibit 14.3). More recently, the market has expanded to include mortgage loans made to credit-impaired borrowers.

The mortgage loans that comprise the HEL ABS sector generally are loans that do not meet the underwriting standards of the government-sponsored agencies (GNMA, FNMA, and FHLMC), which place limits on loan size, loan type, borrower credit rating, and so on. Loans that do not meet the agencies' standards are called nonconforming loans and include jumbo loans (those with high loan amounts) offered to prime borrowers (Alt-A loans) as well as loans to subprime and near-prime borrowers.[5] Cash-out refinance loans, which allow borrowers to access the equity in their homes, dominate the collateral backing most residential ABS transactions. These loans may be fixed rate, adjustable rate (2/28, 3/27, and 5/25 hybrid ARMS), or interest-only (IO) loans.[6] Both adjustable rate and interest-only loans subject the borrower to a payment shock either at the rate reset date or when the loan begins to amortize. Loans to subprime borrowers often contain prepayment penalties, to discourage borrowers from terminating the loan early. Usually prepayment penalties expire with a loan's first rate reset.

The loans that collateralize HEL ABS structures may be fixed rate only, adjustable rate only, or a combination of fixed and adjustable rate loans. These assets can support notes (liabilities) that are fixed rate, adjustable rate, or a combination of fixed and adjustable rates.

Early subprime transactions obtained credit enhancement from monoline insurance companies, in spite of the fact that self-insuring structures were widely used in the prime nonagency market.[7] This was due to limited investor experience in the sector and a relative dearth of collateral performance statistics, such as loan defaults, recoveries and losses. In 1997, the first senior/subordinated structure was introduced. Over time, issuers adopted this structure. As liquidity in the AAA-rated part of the capital structure improved, a market for credit sensitive bonds developed. Now it is not uncommon for transactions to include a spectrum of ratings from AAA to below investment grade. Today the typical structure

Exhibit 14.3 ABS Outstanding ($ Billions as of 2008: Q3)

Home equity	$ 418
Credit card	$ 335
Student loans	$ 244
Auto loans	$ 153
Other	$ 21
Total	$1,171

Source: Securities Industry and Financial Markets Association.

employs a combination of excess interest, overcollateralization, and subordination (see Exhibit 14.4). Loan losses are absorbed in reverse priority order, first by excess interest, then by overcollateralization, and finally via the principal write-down of the subordinated and mezzanine (middle) bonds.

Generally, the underlying mortgage loans are expected to generate more interest income than is required by the liabilities (bonds). Excess interest represents the difference between the weighted average coupon (WAC) rate on the assets and that of the liabilities, net of fees and expenses. When excess interest is positive, it can be used to build a cushion for future potential losses. Overcollateralization (O/C) measures the difference between the total par values of the assets and the bonds. Excess interest can be used to create or maintain a targeted level of overcollateralization.

If delinquency, default, or loss rates are too high, excess interest will be reduced by the amount necessary to compensate for any cash shortfalls experienced by the senior and mezzanine notes. Full or partial repayments and defaults may reduce the amount of excess interest. In general, borrowers with mortgage loans carrying higher interest rates have a greater tendency to prepay. When loans with high interest rates prepay, the weighted average coupon rate on the remaining collateral tends to fall. This is known as WAC drift.

The O/C cushion can accumulate over time, or it can be fully funded when a security is issued. If O/C is accumulated over time, then excess spread is typically used to accelerate the pay-down (principal retirement) of the AAA classes, until a targeted O/C is achieved, usually in the early life of a transaction. If cumulative losses exceed the O/C amount, and if excess spread is insufficient to cover losses in a given period, then subordinated bond investors will incur principal losses.

Like other nonagency mortgage securitizations, subprime transactions employ a shifting interest mechanism that increases the credit enhancement available to the senior bonds. Early in the transaction, principal collections (generated from amortizing, prepaying, maturing, and defaulting loans) and, in some cases, excess interest are paid to the senior bonds only, and the subordinate bonds are locked out from receiving principal.

For the most part, the subordinate bonds are locked out from receiving principal for the first 36 months of a transaction's life, or until the credit enhancement level for the senior bonds has doubled, whichever is later. This point is called the step-down date, referring to a reduction (step-down) in the dollar amount of subordination available as credit enhancement. If loan delinquencies and cumulative losses do not exceed certain levels known as triggers, then at the step-down date,

Exhibit 14.4 HEL ABS Credit Enhancement at Issuance

Class	Rating	% of Notional	% Subordination
Senior	AAA	75	25
	AA	10	15
Mezzanine	A	5	10
Subordinate	BBB	5	5
	BBB–	3	2
Overcollaterization and Excess Interest		2	

Exhibit 14.5 Five-Year Cumulative Loss Rates by Product and Initial Rating, 1993–2005

	CMBS*	RMBS*	HEL ABS*	Global CDO**	ABS***	Global Corporate
Aaa	0.00%	0.03%	0.00%	0.00%	0.03%	0.00%
Aa	0.00%	0.06%	0.00%	0.92%	2.69%	0.08%
A	0.09%	0.34%	0.47%	2.47%	1.31%	0.23%
Baa	0.36%	2.17%	3.42%	10.28%	6.31%	1.24%
Ba	1.40%	3.26%	10.25%	12.60%	21.46%	7.04%
B	9.06%	5.82%	22.44%	27.63%	28.04%	18.61%
Caa	14.88%	19.77%				37.70%
Investment grade	0.14%	0.51%	0.95%	3.73%	1.67%	0.77%
Speculative grade	5.70%	4.42%	14.19%	17.07%	25.71%	26.97%
All Ratings	1.46%	1.00%	2.03%	5.67%	2.69%	6.97%

Source: Moody's *U.S. only; **Excluding CBOs; ***Excluding HEL ABS.

the mezzanine and subordinated classes would receive their pro rata share of principal collections and begin to amortize. If loan delinquencies or cumulative loan losses exceed certain preset triggers, then the mezzanine and subordinated classes would not be entitled to receive any principal, and the issue would convert to a pure sequential pay structure.

The 36-month lockout period is based on historical default experience. Typically, a pool of subprime loans will experience about 60 percent of its total expected cumulative defaults by month 36, with the majority of the defaults occurring between months 24 and 48.

Up until 2007, when subprime stresses became severe, HEL ABS enjoyed a default experience similar to that of corporate bonds. In Exhibit 14.5, we compare the five-year cumulative loss rates for structured credit products and corporate bonds, based on initial credit rating. This data covers the period 1993 to 2005. Note that HEL ABS rated Aaa and Aa outperformed similarly rated corporate bonds. However, HEL ABS rated A and lower underperformed similarly rated corporate bonds. Also note that HEL ABS generally outperformed CDOs, but CMBS rated Ba and higher enjoyed the best overall performance. When these tables are updated to reflect the poor performance of 2006 and 2007 structured credit products, we should expect the loss rates to increase dramatically, especially for HEL ABS and CDOs.

COLLATERALIZED DEBT OBLIGATIONS

The first collateralized debt obligation (CDO) was structured in 1988 and was backed by a collateral pool consisting of high-yield bonds—a collateralized bond obligation, or CBO. Since then, issuers have expanded the menu of reference assets to include leveraged loans (broadly syndicated and middle-market loans), trust-preferred securities, emerging market sovereign debt, asset-backed securities, commercial mortgage-backed securities, commodities, and municipal debt.[8] In Exhibit 14.6, we report statistics on 2007 global CDO issuance. Note that structured finance CDOs (aka ABS CDOs or RESECs, short for resecuritizations) dominate the

Exhibit 14.6 2007 Global CDO Issuance ($ Billions)

Structured finance (ABS CDOs/RESECs)	$255
High-yield loans (CLOs/MM CLOs)	$148
Investment-grade bonds (CBOs)	$ 78
High-yield bonds	$ 2
Other	$ 1
Total	$**486**

Source: Securities Industry and Financial Markets Association.

table, followed by collateralized loan obligations (CLOs), which are populated by high-yield leveraged loans. CDO technology is adaptable, and increasingly issuers are using synthetic assets (loans or securities referenced via credit default swaps) to collateralize an issue.

A CDO is sliced into several notes that represent the issue's liabilities. A note's position in this capital structure determines its priority of claims on the collateral cash flows, which in turn determines its rating. Ratings can range from triple-A for the senior classes to unrated for the first-loss coverage. A CDO only redistributes the total credit risk associated with the pool of assets among the newly created securities. The structure itself neither increases nor reduces the total credit risk associated with the initial pool of assets (see Exhibit 14.7).

CDOs allocate the cash flows of the underlying assets to the notes according to a set of rules, known as a waterfall (see Exhibit 14.8). There are actually two waterfalls, one for interest and one for principal payments. Certain expenses, such as trustee, rating agency, management, and hedge expenses, lie at the top of the waterfall. Then interest cash flows are paid to the notes in sequential order. The notes are paid a fixed spread above a reference rate. Most CDOs pay quarterly interest and reference three-month London Interbank Offered Rate (LIBOR). At the end of the waterfall, the unrated equity or preferred shares receive whatever excess interest remains after the notes have been paid. There are usually two principal waterfalls, one for the reinvestment period and one for the amortization period. During the reinvestment period, principal payments from the collateral are reinvested in new assets. During the amortization period, principal payments are paid either sequentially to the notes in priority order or to all notes simultaneously on a pro rata basis. Again the equity interest is paid last with whatever residual flows exist.

Exhibit 14.7 Typical CLO Structure

Class	Rating	% of Notional	% Subordination	Spread
A	AAA	75%	25%	LIBOR + 50
B	A	9%	16%	LIBOR + 100
C	BBB	6%	10%	LIBOR + 200
D	BB	2%	8%	LIBOR + 500
E	Unrated	8%		Excess

Exhibit 14.8 Typical CDO Interest Cash Flow Waterfall

Interest Proceeds	
Taxes, fees, hedge payments	
Class A interest	
Class A Coverage Tests	
Pass	**Fail**
Class B interest	Class A principal Class B interest when cured
Class B Coverage Tests	
Pass	**Fail**
Class C interest	Class A principal Class B principal Class C interest when cured
Class C Coverage Tests	
Pass	**Fail**
Class D interest	Class A principal Class B principal Class C principal Class D interest when cured
Class D Coverage Tests	
Pass	**Fail**
Subordinated management fee	Class A principal Class B principal Class C principal Class D principal Subordinated management fee when cured
Equity	

Most CDO waterfalls employ additional mechanisms called quality and coverage tests to protect the senior-most investors during periods of stress. The coverage tests compare the interest and principal amounts for the collateral to the interest and principal values of the notes. These tests are applied to each note in the capital structure and compared to a trigger value to determine if cash is to be redirected.

An interest coverage test compares the interest generated by the assets to the interest demanded by the notes that are senior to or commensurate with a particular note. If the interest paid by the collateral is below a targeted level (say, 105 percent of the interest required by the notes senior to or commensurate with a particular tranche), then cash flows are used to pay down the senior most note until the test is in compliance (see Exhibit 14.9). A similar mechanism applies to the principal coverage test, known as an overcollateralization test. Here the par value (or market value) of the collateral is compared to the par value of the notes that are senior to or equal in priority to a particular note. (See Exhibit 14.10.) If the computed par value of the collateral sits below a targeted level (say 104 percent), then interest cash flows are redeployed to pay down the principal of the senior-most notes, until the test is in compliance. Paying down principal of the senior-most notes reduces the numerator in the ratio and lets the computed value converge to the trigger.[9]

The OC and IC trigger values will vary from one issue to the next and will be a function of the assets that collateralize the CDO as well as the amount of

Exhibit 14.9 Interest Coverage Test

$$\text{IC ratio for tranche C} = \frac{\text{Collateral scheduled interest}}{\text{Sum of interest required on Tranches C, B, and A}}$$

Example: \$1M collateral; 7.5% WAC; no defaults; LIBOR = 4%*

Class	IC Ratio	Trigger	Pass
A	(\$1M × 7.5%)/(\$750K × 4.5%) = 222%	125%	Y
B	(\$1M × 7.5%)/((\$750K × 4.5%) + (\$90K × 5%)) = 196%	115%	Y
C	(\$1M × 7.5%)/((\$750K × 4.5%) + (\$90K × 5%) + (\$60K × 6%)) = 179%	105%	Y

*See Exhibit 14.7 for spreads.

subordination. Generally, par value triggers are more stable than market value triggers, which are commonly used in market value CDOs. These are a small sector of the CDO universe (5 percent or less), and these structures tend to resemble actively managed portfolios. Par value triggers may apply what are called haircuts to certain assets, should their ratings fall below a minimum level. These haircuts would discount the par amount by a given percentage, thus reducing the coverage.

CDOs may be static or managed. With a static CDO, the reference assets do not change over time. With a managed CDO (more common), a manager trades the portfolio of assets. Some discretionary trading is allowed, typically 10 to 20 percent of the net asset value per year. Additionally, managers will reinvest assets that prepay, mature, or default during the reinvestment period. The reinvestment period varies from 3 to 7 years. Generally, the notes of a CDO have average lives from 7 to 10 years, although their legal, final maturities may be 30 or 40 years.

CDOs also employ quality tests with place limits on the asset portfolio. A weighted average rating factor (WARF) test places restrictions on the average credit rating of the assets. Other quality tests place limits on the total allocations to a particular rating, issuer, type of security, industry, sector, and so on.

Managers typically take a stake in the equity tranches of CDOs. As equity investors, they have an incentive to maximize residual cash flows. Obviously, this encourages them to increase risk. However, they also need to balance the interests of senior investors. If they err on the side of aggression, they may suffer reputation losses. If they err on the side of caution, they may generate lackluster returns for

Exhibit 14.10 Overcollateralization Coverage Test

$$\text{O/C ratio for tranche C} = \frac{\text{Collateral par value}}{\text{Sum of par values on tranches C, B, and A}}$$

Example: \$1M collateral; no defaults

Class	O/C Ratio	Trigger	Pass
A	\$1M/\$750K = 133%	111%	Y
B	\$1M/(\$750K + \$90K) = 119%	108%	Y
C	\$1M/(\$750K + \$90K + \$60K) = 111%	104%	Y

themselves and other equity investors. In general, there are mechanisms to balance these competing interests. Often manager fees are split into senior and subordinate fees. Managers may also hold senior notes to better align their performance with senior investors. The largest CDO managers are asset management firms, some of which are affiliated with investment and commercial banks. These include TCW, Duke Funding, BlackRock, Highland Capital, and Credit Suisse Asset Management.

The default performance of CDO notes historically has been strong, despite weakness in CBOs and early-vintage structured finance CDOs. Five-year cumulative loss rates for the period 1993 to 2005 are presented in Exhibit 14.5. ABS CDOs have come under severe stress, and the rating agencies have been accused of misrating this product.

COMMERCIAL MORTGAGE-BACKED SECURITIES

Commercial mortgage-backed securities are structured bonds backed by commercial real estate mortgage loans. Though the first transactions were privately placed and issued in the late 1980s, it was not until the establishment of the Resolution Trust Corporation in 1989 that the market really took off. At nearly $800 billion, CMBS outstanding currently represent approximately 25 percent of all commercial mortgage loans outstanding in the United States.

Commercial mortgage loans reference retail (shopping or strip malls), office, multifamily, hotel, industrial, storage, and healthcare properties. Generally the loan amounts on these properties are large (between $1 million and $1.5 billion), and the loan terms are fixed rate, incorporating balloon and prepayment penalty features.

Investors rely on two metrics to evaluate the loans in CMBS transactions. The debt service coverage ratio (DSCR) measures a property's net operating income relative to the mortgage payment required on the underlying loan. A DSCR above 150 percent is generally regarded as very good. The loan to value (LTV) ratio is related to a property's net cash flow and its capitalization rate (essentially the return an investor would require if the property were purchased with cash). An LTV of 65 percent or lower is desirable.

CMBS investors benefit from loan diversification across sectors and geographical regions. Moody's uses a Herfindahl index to score loan concentrations.[10] For example, if a pool of 50 loans has a Herfindahl index of 25, this indicates that the pool has an effective diversity of 25 loans.

CMBS resemble residential mortgage securitizations, but CMBS use sequential pay bond structures and offer investors greater call protection thanks to the use of lockout terms (for two to five years), yield maintenance penalties, and defeasance. With defeasance, the borrower purchases a basket of Treasury securities that replicate the scheduled cash flows on a loan, thus insuring against a prepayment.

Because of increased investor interest and excellent near-term performance, CMBS subordination levels have been falling since the mid-1990s. In a typical current transaction, the senior AAA class would enjoy 12 percent subordination, down from more than 30 percent in 1995. Unlike some other fixed income assets, CMBS are priced at spread-to-swap rates. In the example in Exhibit 14.11, the AAA class might earn 60 basis points above swap rates.

Exhibit 14.11 CMBS Classes and Subordination Levels

Class	Rating	Type	Subordination
A	AAA	Fixed spread over swap rates	12%–34%
B	AA	Fixed spread over swap rates	10%–28%
C	A	Fixed spread over swap rates	7.5%–22%
D	BBB	Fixed spread over swap rates	4.5%–15%
E	BB	Fixed spread over swap rates	3%–7%
F	B	Fixed spread over swap rates	2%–3%
G	Unrated	Excess spread	0
X-P		PAC interest only	
X-C		Support interest only	

Like residential mortgage securitizations, CMBS transactions often include interest-only bonds. In Exhibit 14.11, the X-P and X-C classes are interest-only bonds. The cash flow to pay these bonds comes from the excess interest available in the structure. The X-P is called a planned amortization class (PAC) IO bond and the X-C is the support or levered IO bond. PAC bonds have very stable cash flows, which are essentially guaranteed by simultaneous-pay support bonds that absorb prepayment risks. In CMBS transactions, PAC bonds usually are structured to withstand high prepayment rates (e.g., 6 percent constant annual default rates).

CMBS have enjoyed a very strong record of performance, at least when it comes to rating upgrades and downgrades. In general, CMBS have recorded a better rating transition profile than corporate bonds at all ratings. (See Exhibit 14.5.)

ENDNOTES

1. Three companies participate actively in the credit rating process for structured credit products: Standard and Poor's, Moody's, and Fitch.
2. Government guaranteed mortgage loans, such as FHA or VA loans, are typically securitized by GNMA. FNMA and FHLMC securitize loans that conform to their underwriting standards.
3. Until recently, only GNMA enjoyed a true Federal guarantee. Before September 2008, when FNMA and FHLMC were placed in conservatorship, their securities traded with an implicit government guarantee.
4. Many fixed-income securities (e.g., Treasury notes and non-callable corporate bonds) possess positive convexity. With positive convexity, a unit decline in interest rates produces ever larger bond price gains as rates approach zero. Residential mortgage loans suffer from negative convexity: as interest rates fall, the price appreciation on a mortgage loan is capped due to the increased likelihood that the borrower will refinance and prepay the loan at par.
5. Prime borrowers are those with FICO (Fair Isaac & Co) scores above 679; subprime borrowers are those with FICO scores below 621.
6. Adjustable rate mortgage loans have coupons that float above a benchmark interest rate, such as LIBOR. An initial teaser rate may be used to provide lower monthly borrower payments initially. The time to first rate reset varies from 2 years (so called 2/28 loans) to 5 years (5/25 loans).
7. The largest mono-line insurers are MBIA and AMBAC.

8. Leveraged loans are bank loans extended to non-investment grade borrowers. The loans are collateralized by real assets and generally enjoy higher recoveries than unsecured corporate bonds. Leveraged loans can be broadly syndicated (issued to medium size companies and traded by a consortium of large commercial and investment banks), or so called "middle market " loans made to smaller companies (those with EBITDA of $50 million or less).
9. Trigger-less CDOs, which are usually populated by HEL ABS notes rated A or lower, dispense with most I/C and O/C tests and consequently afford senior note investors fewer protections.
10. The Herfindahl index is a measure of industry concentration. It is computed as the sum of the squares of an industry's individual firm market shares.

REFERENCES

"Understanding Residential ABS: A Comprehensive Guide to Collateral, Structure and Related Credit Derivative Markets," Glenn Schultz, *Wachovia Capital Markets, LLC*, February 14, 2007.

"Deconstructing CDOs: An Updated Manual for an Evolving Structure as CDOs Hit the Trillion-Dollar Mark," Brian McManus, Dave Preston, Steven Todd, and Anik Ray, *Wachovia Capital Markets, LLC*, May 23, 2007.

"CMBS Primer: An Introduction to Commercial Mortgage-Backed Securities," Brian Lancaster, and Anthony Butler, *Wachovia Capital Markets, LLC*, April 17, 2007.

Hirschman, Albert. 1964. "The Paternity of an Index," *American Economic Review* 54: 761–762.

"Moody's Default and Loss Rates of Structured Finance Securities: 1993–2006," *Moody's Investors Service*, April, 2007.

"Moody's Corporate Default and Recovery Rates, 1920–2006," *Moody's Investors Service*, February, 2007.

The Securities Industry and Financial Markets Association, www.sifma.org.

ABOUT THE AUTHOR

Steven Todd is an associate professor of finance at Loyola University Chicago. Dr. Todd completed his Ph.D. in finance and business economics at the University of Washington; he also holds a B.S. degree in operations research from Cornell University. Dr. Todd teaches graduate and undergraduate courses in corporate finance, investments, options, derivatives, credit risk management, international finance, portfolio management, and investment banking. His research interests include asset pricing, financial markets and institutions, and corporate governance. He has published papers on securitization, mutual fund performance measurement, managerial compensation, and equity analysts in several journals, including *Real Estate Economics*, *Journal of Business*, *Journal of Corporate Finance*, and *Financial Markets*, *Institutions and Instruments*. Prior to his academic career, Dr. Todd worked on Wall Street as an analyst, investment banker and trader at Merrill Lynch and UBS Securities. Recently, Dr. Todd served as a director in the CDO Research Group at Wachovia Securities.

CHAPTER 15

Executive Stock Options

ROBERT W. KOLB
Professor of Finance and Frank W. Considine Chair of Applied Ethics,
Loyola University Chicago

INTRODUCTION

The nature and amount of executive compensation is an enduring flashpoint of public disagreement and an issue of great moment in corporate governance. While many object to the system of executive compensation in the United States based on the total size of pay packages, the use of executive stock options is especially controversial. The outrage over executive stock options is matched only by the lack of understanding of how these options are supposed to work. This chapter introduces the basics of executive stock options, explains the rationale for their widespread use, considers some of the more cogent criticisms of these options, and points the way to a more complete understanding of these instruments based on a voluminous literature that continues to grow.

BASIC FEATURES OF EXECUTIVE STOCK OPTIONS

A typical pay package for a leading executive in a public company is composed of four elements: salary, bonus, stock options, and retirement or pension. (In addition, an executive has a significant opportunity to consume perquisites, such as company planes, apartments, and so on, but we do not address those in this chapter.) Under current U.S. law, executive compensation in excess of $1 million per year is not a deductible corporate expense *unless* the additional compensation is performance based in some way. Thus, most firms limit salaries to $1 million, even for the highest-level executives. Nonetheless, the pay packages can be worth many millions, with the additional amounts being made up of bonus monies and executive stock options (ESOs). Of these other elements, ESOs are quite frequently the largest; in fact, they may exceed 50 percent of the total value of pay.

A typical ESO might be granted to an executive with an initial term to expiration of 10 years. Until recently, tax laws in the United States provided that an ESO issued with a strike price equal to the current stock price of the firm generated no expense for accounting purposes. Thus, for many years, virtually all ESOs were issued with a strike price equal to the exercise price. (Now current law requires recognition of the value of an options grant as an expense in the period in which it is granted.) ESOs are nontradable, and the holder can capture the value of these options through only exercise or through an exchange with the issuing firm. ESOs also often carry a vesting requirement of four years, such that an employee who leaves the firm during the vesting period forfeits the option upon termination. If the holder of a vested option leaves the firm, the option must be exercised if it is in-the-money or forfeited if it is out-of-the-money. The typical ESO is a kind of Bermudan options: halfway between a European- and an American-style option in the sense that it is better than a European option in not permitting exercise only at expiration but not as good as an American option, which can be exercised at any time. This basic description hold for a very large percentage of ESOs, but there are some variations.

Some ESOs are granted with a strike price that is indexed to some other variable, such as the rate of change in a stock index or to the firm's cost of capital, such that the exercise price rises over time. This practice is rare, as the no-expensing feature of earlier tax law required that the options have a strike price equal to the current stock price and an indexed strike price violated that provision. Many ESOs also have a reload provision, such that an executive who exercises an option receives replacement options with a new strike price equal to the stock price at the time of exercise. The number of replacement options is often equal to the number of options exercised, but that is not always the case. The time to expiration of the new option can be the same as the remaining term to expiration as the original option, but it sometimes can be longer. As a final feature, some firms allow a resetting of the strike price in some circumstances. When an option is issued at one strike price (that would typically match the stock price at the time of issuance), but the firm's stock falls in value, the firm may reset the strike price of the existing options to a new lower strike price equal to the current lower stock price. This essentially amounts to a canceling of the existing option and an issuance of new options at the current lower stock price.

Rationales for ESOs

As noted, ESOs have long been a feature of the executive compensation landscape, but the rise of ESOs to real prominence dates from 1990 and the publication of an article by Jensen and Murphy. In this article, the authors decry recent trends in corporate pay, noting that chief executive officers (CEOs) in general have little direct financial stake in the fortunes of their firms. Surveying the period from 1974 to 1998, they note that CEO stock holdings as a percentage of corporate value has been declining, and that "a $1,000 change in corporate value corresponds to a change of just 6.7 cents in salary and bonus over two years" (p. 139). They explicitly recommend that "cash compensation should be structured to provide big rewards for outstanding performance and meaningful penalties for poor performance" and that such generous rewards should be coupled with high demands for performance including making real the "threat of dismissal" (p. 141, 142).

The purpose in this recommended change in the structure of compensation was to bring compensation in line with the agency theory of the firm. According to this view of the firm, the CEO is an agent of the shareholders, who are the principals in the firm. The principals contract with their agent to operate the firm on behalf of shareholder interests. Yet there is always a divergence between the desires of the agent and those of the principal. Therefore, the agent's employment contract should be written in a way to mitigate the "incentive incompatibility" between the principal and agent and to strive for "incentive alignment." In other words, the ideal CEO contract should induce the CEO to operate the firm just as the shareholders would if they were on the scene and possessed of the CEO's expertise. This understanding of the role of equity-based executive compensation is known as the incentive alignment theory or the optimal contracting approach. (This theory recognizes that a perfect contract is impossible, so there is always some remaining incentive incompatibility, a problem that is addressed later in the chapter.) Other aspects of the incentive alignment approach stress executive attraction and retention, and the use of ESOs to attract the "right" kind of executives (i.e., performance-oriented executives willing to undertake risky high net present value projects).

Shareholders presumably want an increasing stock price, so one way of aligning the CEO's incentives with the shareholders is by paying the CEO with an equity stake in the firm. This can be achieved in two main ways: by giving the CEO restricted shares in the firm or by granting ESOs. Because of the inherent leverage of options, paying the CEO with $1 of options rather than $1 of shares makes the CEO's wealth more responsive to the firm's share price for a given compensation cost to the firm.

In addition to incentive alignment, paying employees with shares can preserve scarce cash, especially in start-up firms. In the technology wave of the 1990s and the ensuing dot-com bubble, firms moved strongly toward the compensation structure advocated by Jensen and Murphy (1990). Equity-based compensation, especially in the form of ESOs, became an extremely important part of executive pay packages. This movement is understood by incentive alignment theorists as a happy outcome, with today's pay packages being granted by boards to executives to make them more effective agents of the shareholders.

There is, however, a competing explanation for the rise of ESOs. The optimal contracting approach requires that the executive's pay package be set in such a way that the CEO behaves as the shareholders wish. But what if the CEO has an important voice in determining his or her own pay? The managerial power approach maintains that this is exactly the case: The CEO and other top executives exercise influence over their own pay packages in a way that results in a markedly suboptimal contract, and that has the result of illegitimately transferring wealth from shareholders to executives.

This theory is most strongly associated with Lucian Arye Bebchuk and Jesse M. Fried (2003, 2005). In the Bebchuk and Fried analysis, executives enrich themselves at shareholder expense in a variety of ways. For example, CEOs sometimes even serve on the executive compensation committee of the board, the very group charged with setting CEO pay. Less blatantly, CEOs often have considerable influence over board appointments. Also, according to the Bebchuk and Fried analysis, CEOs and directors form a "club" in which a mutual identification of interests and a kind of mutual "back-scratching" lead to excessive compensation. Bebchuk and Fried do not altogether deny that incentive alignment also occurs, but they maintain that the optimal contracting approach provides merely a part of the story that must be supplemented by their managerial power perspective.

Bebchuk and Fried argue that there are limits to how far this shareholder abuse can go. Pressed too far, grossly excessive CEO pay packages would lead to public and shareholder outrage, and this "outrage cost" provides a limiting factor for executive pay packages. Further, the desire to expand executive compensation and to avoid outrage costs leads to various efforts to camouflage the magnitude of executive pay. According to Bebchuk and Fried, this occurs in several ways, including excessively generous pension and retirement plans. For our purposes, the Bebchuk and Fried account also stresses that one way of camouflaging pay is by using ESOs. As the next section explains, the complex features of ESOs makes their value less than transparent, and for Bebchuk and Fried, this lack of transparency makes ESOs a powerful vehicle for camouflaging executive pay. There is a vast empirical literature investigating both the incentive alignment theory and the managerial power approach, and the last section of this chapter titled "Executive Stock Options and Incentives" touches on some representative studies.[1]

Pricing of Executive Stock Options

Compared to typical exchange-traded options, executive stock options have a number of features that make them difficult to value. These include vesting, non-tradability, the possibility of forfeit, price resetting and reload provisions, and the tendency for holders of these options to exercise options in a suboptimal way, thereby discarding a portion of the option's value.

Black and Scholes achieved their famous pricing result by assuming away market frictions, thereby allowing investors to trade fractional shares continuously and without transaction costs. This meant that it was possible to create a portfolio of an option and the underlying stock that was risk free and to adjust the portfolio continuously in a manner that kept it free of risk. This strong condition is entirely inapplicable to executive stock options that cannot be traded. Further, executives who hold ESOs are not permitted to hedge their option positions by selling the firm's shares short. (Permitting executives to sell the firm's shares short would absolutely frustrate the incentive alignment dimensions of equity compensation.)

These restrictions on the behavior of holders of ESOs drives a conceptual wedge between the cost of the option to the firm and the value of the option to the holder of the ESO. Usually the valuation of the ESO refers to the present value of the expected costs incurred by the firm in issuing the option. While there are some closed-form and analytical models of ESO valuation, most valuation tends to proceed by exploiting lattice methods.[2]

From the point of view of the issuing firm, most of the features of ESOs make their cost less than that of the otherwise similar plain American option. For example, a vesting period moves the option away from being an American option and toward being a European option, with a corresponding reduction in price, at least for ESOs on stocks that pay dividends. If the holder of an ESO leaves the firm, he or she must terminate the option. If the option is unvested, the departing executive forfeits the value of the option; if the option is vested but out-of-the-money when the executive leaves the firm, the actual cost to the firm is also zero. At the time the ESO is granted, the potential for termination and forfeiture reduce the expected cost of issuing the option. The fact that the options are nontradable also affects the cost of these options to the granting firm. After vesting, with the holder of the ESO continuing with the firm and the option in-the-money, holders frequently exercise their options well in advance of expiration. In the analysis of tradable options and ignoring dividends, exercise before termination is always a mistake. Exercise captures the exercise value of the option but discards the option's time value, so it is always better to sell the option rather than exercise it.

But an ESO is not tradable, and there are many reasons why ESO holders might rationally exercise the option before expiration. For example, an executive with an ESO might want immediate access to the value of the option for other purposes. Alternatively, the executive might want to reduce her exposure to the risk of the firm. In this case, exercising the option restores a desired portfolio balance. Whatever the reason for early and suboptimal exercise, the issuing firm benefits, because part of the full value of the option is discarded and the actual cost to the firm is reduced. At the time of the issuance, the firm can anticipate these suboptimal exercises and can realize that the expected cost of issuing the option is less than the cost of issuing an otherwise similar fully tradable option.

These various features of ESOs are probably best handled by a lattice model, which has great flexibility in including the impact of vesting, the possibility of forfeiture, and the possibility of suboptimal exercise. But all of these models examine the option from the point of view of the firm to provide an answer to the question: What is the cost of issuing the option to the firm?

There is a separate and parallel question, however: What is the value of the option to the recipient? Because the holder cannot hedge the option and cannot trade it, it is quite possible that the value of the option to the holder could be less than the firm's cost of issuing the option. A number of models attempt to capture that prospective differential by modeling the utility function of the executive. These models almost invariably use a power utility function in which utility is a function of the executive's level of wealth at time t, W_t, and the executive's coefficient of relative risk aversion, λ. Thus in equation 15.1 utility, U_t, is expressed as

$$U_t = \frac{{W_t}^{1-\lambda}}{1-\lambda} \tag{15.1}$$

for a representative individual estimates of λ range from 2 to 9 and even beyond. Even assuming that an individual's utility function is reasonable described by a power function, any individual's level of risk aversion is likely to remain a mystery to all. (This would include the executive herself. After all, what is your λ?)

Nonetheless, these models emphasize the difference between the cost to the firm and the value to the employee, and they provide an interesting heuristic.[3] The underlying conception is that the executive is risk averse. Because part of the executive's compensation is provided as an ESO, which cannot be hedged or traded, the executive holds a less than optimally balanced portfolio with excessive commitment to the firm's shares. After all, the executive's human capital is committed to the firm, and ESOs tie the executive's wealth even more fully to the fate of the firm. As a consequence, these utility-based models uniformly show that the value of the option to the executive is less than the cost of the option to the issuing firm. To make this explicit, assume that the firm estimates that the cost of granting an option is \$15, taking into account the expected cash flows from granting the option after considering the influence of possible forfeits and early exercise. If the value to the executive is less than the cost to the firm, that means that the executive would willingly surrender the option for less than the \$15 cost to the firm. This means that ESO compensation is inefficient—that is, the cost of the option to the issuing firm exceeds the value received by the executive.

All utility-based models show this gap between cost to the firm and value to the employee for risk-averse executives. (If $\lambda = 0$, then the executive is risk neutral, and the cost to the firm and value to the employee are the same.) However, these models leave out one very important and central dimension of ESOs.

The legitimate rationale for granting ESOs on the optimal contracting approach is to alter the executive's incentives and to change her behavior in a manner that leads to an increase in firm value over what it would have been had the firm not provided the incentive of ESOs. Yet the models value the options without considering the ability of executives to influence the price of the firm's shares. If the executive positively increases the value of the firm due to the incentives received, then the cost to the firm, the payoff to the executive, and the utility received by the executive are all higher than the typical models reflect. This point is made quite forcefully by Tung-Hsiao Yang and Don M. Chance (2008), who develop a model in which the value of the option to the executive can be higher than the cost of issuing the option by the firm, all conditional on the executive's being able to influence the course of the firm's stock value.

Nonetheless, the dominant view in the literature portrays a cost to the firm substantially less than the value of an otherwise-similar American option, due to the special features of ESOs that we have been exploring. Further, the utility-based analyses almost universally portray the value received by the executive as significantly lower than the cost to the firm. That is, the consensus view regards ESOs as a form of inefficient compensation, in which the value placed on these options by the executive is less than the cost to the firm of issuing the option. One may well wonder, then, why firms would use such a form of compensation? We have seen that the managerial power hypothesis would insist that part of the explanation lies in executives' ability to influence compensation and to receive excessive compensation in the camouflaged form of ESOs. Even if ESOs are an inefficient means of compensation, the firm still could win if the ESOs provide incentives that result in an increase in firm value that more than offsets the inefficiency cost of using ESOs in executive compensation. In short, the optimal contracting approach maintains that a legitimate explanation for using inefficient ESOs lies in the role of ESOs in creating incentives for superior firm performance.

Executive Stock Options and Incentives

ESOs can be powerful incentive devices, and their effectiveness for both good and ill is well documented in a vast literature. Because the incentive effect of ESOs is so well studied, we will consider only some representative studies for effective incentivization through ESOs and some of the more striking perverse incentives provided by ESOs. A consideration of these incentive effects will shed some light on the controversy between the optimal contracting and market power understandings of ESOs.

The basic purpose of granting ESOs according to the optimal contracting approach is to bring the executive's incentives into alignment with those of the firm in general and shareholders in particular. The theory does not require any particular conception of the executive's preexisting incentives, but the typical scenario of most optimal contracting theorists portrays a risk-averse CEO, who wants to quietly manage the firm in a low-risk manner while he or she enjoys a peaceful life with a high salary and the consumption of rich perquisites. Alternatively, some studies assert that before the emphasis on ESOs, many managers wanted to increase firm size or create empires, as compensation was related more to the scale of the firm than to the firm's performance.[4]

If ESO plans are beneficial to shareholders, one would expect a positive stock price reaction to the announcement of a new plan, and this seems to be the case (Kato, Lemmon, Luo, and Schallheim, 2005; Morgan and Poulsen, 2001). In the aftermath of the subprime and credit crisis of 2007 and 2008, it may seem strange to think it is a good idea to incentivize CEOs to increase risk, but if one conceives of CEOs as risk-averse managers who sacrifice firm value to avoid risk, then increasing the CEO's incentive to undertake risky positive net present value projects would be a good idea. Evidence suggest that ESO plans are associated with greater risk taking (Coles, Daniel, and Naveen, 2006; Rajgopal and Shevlin, 2002). Institutional investors generally are considered to be savvy investors, and their investment in firms is generally regarded as a sign of good corporate governance. There is a strong tendency for such institutional investors to gravitate toward firms with ESO plans (Hartzell and Starks, 2003). In addition, there is evidence that CEOs with ESO plans make decisions that enhance shareholder value, including more investment in research and development, focusing on fewer lines of business, and undertaking layoffs that lead to stock price increases (Brookman, Chang, and Rennie, 2007; Coles et al., 2006).

In contrast to these positive indicators, it is also clear that ESOs lead to some bad management practices and even criminal behavior. It is well known that missing earnings targets adversely affect share prices and also ESO values. There is very strong evidence that earnings misreporting, management, and manipulation are associated with ESOs.[5]

ESOs can also lead managers to make poor management decisions. For example, ESOs are not dividend protected, and dividends reduce share prices and lower the values of ESOs. Firms with ESO plans substitute share repurchases for dividends. This substitution of repurchases for dividends is not necessarily bad, but it shows that ESOs can distort managerial policies (Fenn and Liang, 2001). In general, an option is more valuable if the underlying stock is riskier. An executive with sharply underwater (out-of-the-money) options can have an incentive to take wildly risky projects, even if they are negative net present value projects, just to try to make the ESOs pay off (Rogers, 2005).

The existence and magnitude is also associated with the frequency of shareholder suits and fraud charges (Denis, Hanouna, and Sarin, 2005; Peng and Röell, 2006). The existence of ESOs has led to criminal actions that transfer wealth from shareholders to executives. The most famous instance of this is option grant-date backdating—reporting a grant date that is earlier than the actual grant date, when the stock price on the falsely reported grant date is below the stock price on the actual grant date. As an example, consider a rising stock price environment with a stock price of \$50 on the grant date. In backdating, the firm would report the grant date as an earlier date when the stock price was lower, say \$40. This immediately puts the option in-the-money and surreptitiously transfers wealth from shareholders to the executive. In many instances, this backdating has been accompanied by the falsification of documents, intentional misreporting of accounting results, and the evasion of taxes. A number of CEOs have gone to prison for backdating.[6]

CONCLUSION

As we have seen, ESOs have proven to provide powerful incentives—perhaps too powerful. And they have proven to provide incentives for both value-enhancing management decisions as well as serious misbehavior and criminal conduct. There is considerable evidence that ESOs perform as the optimal contracting approach stipulates: ESOs can provide incentives for executives to manage more strenuously in the pursuit of shareholder value. But it seems equally clear that the optimal contracting approach is not the entire story. Rather, the evidence of executive self-dealing and rent extraction is equally compelling. Perhaps not surprisingly the use of ESOs to align incentives works well only in an environment of highly effective corporate governance and strong monitoring by a well-functioning board of directors.

In the aftermath of the financial crises of 2007–2008, it seems quite likely that the immediate future will see more conservative business practices with much lower leverage ratios and a greater emphasis on capital adequacy, accompanied by more prudent, or even cautious, business management. The strong backlash against executive pay may well lead to more effective restrictions on the magnitude of executive pay. If pay packages are restrained, it is quite likely that ESOs will diminish in absolute terms, and they may well form a smaller portion of executive compensation. Nonetheless, ESOs will continue to be an important feature of the executive compensation landscape.

ENDNOTES

1. For a more complete review of this literature and an extensive bibliography, see Kolb (forthcoming).
2. Cvitanić, Wiener, and Zapatero (2008) present a quite robust and complex analytical model, while Ammann and Seiz (2004) compare a variety of models of different types (closed form, analytical, and lattice). Hull and White (2004) present a lattice model that takes account of early exercise, while Sircar and Xiong (2007) present a lattice model that takes account of many of the complex features of ESOs.
3. Two of the earliest utility-based models were Kulatilaka and Marcus (1994) and Rubinstein (1995). Chance and Yang (2005) provide a more recent and more full-featured model.
4. For a discussion of these two managerial habits, see Jensen and Murphy (1990) and Bertrand and Mullainathan (2003).
5. Just as a sample of this voluminous literature, see Bartov and Mohanram (2004); Bergstresser and Philippon (2006); Burns and Kedia (2006).
6. The seminal backdating articles are: Heron and Lie (2007) and A. Heron, Lie, and Perry (2007). In addition to grant-date backdating, another type of backdating is exercise-date backdating, in which the date on which an employee stock option is exercised is falsely reported. The idea here is to falsely report an exercise on a date when the stock price was low, then when the stock is held, the strategy gains beneficial tax treatment. See Cicero (2007).

REFERENCES

Ammann, M., and R. Seiz. 2004. "Valuing Employee Stock Options: Does the Model Matter?" *Financial Analysts Journal* 60, No. 5: 21–37.

Bartov, E., and P. Mohanram. 2004. "Private Information, Earnings Manipulations, and Executive Stock-Option Exercises," *Accounting Review* 79, No. 4: 889.

Bebchuk, L. A., and J. M. Fried. 2003. "Executive Compensation as an Agency Problem," *Journal of Economic Perspectives* 17, No. 3: 71–92.

Bebchuk, L. A., and J. M. Fried. 2004. "Stealth Compensation via Retirement Benefits," Working paper.

Bebchuk, L. A., and J. M. Fried. 2005. "Executive Compensation at Fannie Mae: A Case Study of Perverse Incentives, Nonperformance Pay, and Camouflage," *Journal of Corporation Law* 30: 807–822.

Bebchuk, L. A., and J. M. Fried. 2005. "Pay Without Performance: Overview of the Issues," *Journal of Corporation Law* (Summer): 647–673.

Bebchuk, L. A., and Y. Grinstein. 2005. "The Growth of Executive Pay," *Oxford Review of Economic Policy* 2: 283–303.

Bergstresser, D., and T. Philippon. 2006. "CEO Incentives and Earnings Management," *Journal of Financial Economics* 80, No. 3: 511–529.

Bertrand, M., and S. Mullainathan. 2003. "Enjoying the Quiet Life? Corporate Governance and Managerial Preferences," *Journal of Political Economy* 111: 1043–1075.

Brookman, J. T., S. Chang, and C. G. Rennie. 2007. "CEO Cash and Stock-Based Compensation Changes, Layoff Decisions, and Shareholder Value," *Financial Review* 42: 99–119.

Burns, N., and S. Kedia. 2006. "The Impact of Performance-Based Compensation on Misreporting," *Journal of Financial Economics* 79, No. 1: 35–67.

Chance, D. M., and T.-H. Yang. 2005. "The Utility-Based Valuation and Cost of Executive Stock Options in a Binomial Framework: Issues and Methodology," *Journal of Derivatives Accounting* 2, No. 2: 165–188.

Cicero, D. C. 2007. "Strategic Timing and Backdating of Executive Stock Option Exercises: Before and After the Sarbanes-Oxley Act." U.S. Securities and Exchange Commission, February 4.

Coles, J. L., N. D. Daniel, and L. Naveen. 2006. "Managerial Incentives and Risk Taking," *Journal of Financial Economics* 79: 431–468.

Cvitanić, J., Z. Wiener, and F. Zapatero. 2008. "Analytic Pricing of Employee Stock Options," *Review of Financial Studies* 21, No. 2: 683–724.

Denis, D. J., P. Hanouna, and A. Sarin. 2005. "Is There a Dark Side to Incentive Compensation?" Working paper (March).

Fenn, G. W., and N. Liang. 2001. "Corporate Payout Policy and Managerial Stock Incentives," *Journal of Financial Economics* 60, No. 1: 45–72.

Hartzell, J. C., and L. T. Starks. 2003. "Institutional Investors and Executive Compensation," *Journal of Finance* 58, No. 6: 2351–2374.

Heron, R. A., and E. Lie. 2007. "Does Backdating Explain the Stock Price Pattern Around Executive Stock Option Grants?" *Journal of Financial Economics* 83, No. 2: 271.

Heron, R. A., E. Lie, and T. Perry. 2007. "On the Use (and Abuse) of Stock Option Grants," *Financial Analysts Journal* 63, No. 3: 17–27.

Hull, J., and A. White. 2004. "How to Value Employee Stock Options," *Financial Analysts Journal* 60, No. 1 (January–February): 114–119.

Jensen, M. C., and K. J. Murphy. 1990. "CEO Incentives: It's Not How Much You Pay, But How," *Harvard Business Review* (May–June): 138–149.

Kato, H. K., M. Lemmon, M. Luo, and J. Schallheim. 2005. "An Empirical Examination of the Costs and Benefits of Executive Stock Options: Evidence from Japan," *Journal of Financial Economics* 78, No. 2: 435–461.

Kolb, R. W. 2010 (forthcoming). *Executive Stock Options: Financial, Social, and Ethical Issues* Hoboken, NJ: John Wiley & Sons.

Kulatilaka, N., and A. J. Marcus. 1994. "Valuing Employee Stock Options," *Financial Analysts Journal* 50, *No. 6* (November–December): 46–56.

Lambert, R. A., W. N. Lanen, and D. F. Larcker. 1989. "Executive Stock Option Plans and Corporate Dividend Policy," *Journal of Financial and Quantitative Analysis* 24, No. 4 (December): 409–425.

Marquardt, C. A. 2002. "The Cost of Employee Stock Option Grants: An Empirical Analysis," *Journal of Accounting Research* 40, No. 4: 1191–1217.

Morgan, A., and A. Poulsen. 2001. "Linking Pay to Performance-Compensation Proposals in the S&P 500," *Journal of Financial Economics* 62: 489–523.

Peng, L., and A. Röell. 2006. "Executive Pay and Shareholder Litigation," Working paper (August).

Rajgopal, S., and T. Shevlin. 2002. "Empirical Evidence on the Relation between Stock Option Compensation and Risk Taking," *Journal of Accounting & Economics* 33, No. 2: 145–171.

Rogers, D. A. 2005. "Managerial Risk-Taking Incentives and Executive Stock Option Repricing: A Study of U.S. Casino Executives," *Financial Management* 34, No. 1 (Spring): 95–121.

Rubinstein, Mark. 1995. "On the Accounting Valuation of Employee Stock Options," *Journal of Derivatives* (Fall): 8–24.

Sircar, R., and W. Xiong. 2007. "A General Framework for Evaluating Executive Stock Options," *Journal of Economic Dynamics & Control* 31: 2317–2349.

Yang, T.-H., and D. M. Chance. 2008. "The Valuation of Executive Stock Options when Executives Can Influence the Payoffs," Working paper (January).

ABOUT THE AUTHOR

Robert W. Kolb is professor of finance at Loyola University of Chicago, where he also holds the Frank W. Considine Chair of Applied Ethics. Dr. Kolb received two PhDs from the University of North Carolina at Chapel Hill in philosophy and finance (philosophy 1974, finance 1978). Previously he held full-time academic positions at the University of Florida, Emory University, the University of Miami, and the University of Colorado. Over his career, he has published more than 50 academic research articles and more than 20 books, most focusing on financial derivatives and their applications to risk management. In 1990, he founded Kolb Publishing Company to publish finance and economics university texts. He sold the firm to Blackwell Publishers of Oxford, England, in 1995.

CHAPTER 16

Emerging Derivative Instruments

STEVE SWIDLER
J. Stanley Mackin Professor of Finance, Auburn University

For nearly 300 years, derivatives trading largely entailed the future delivery of an agricultural commodity. Two early examples of exchange traded derivatives include forward contracting of tulip bulbs on the Royal Exchange in London around 1637 and a standardized rice contract for future delivery that traded on the Yodoya rice market in Osaka, Japan, approximately 13 years later (Chance, 1998). In the United States, the creation of the Chicago Board of Trade in 1848 led to the buying and selling of grain futures, and later in the century, eggs and butter were the underlying assets for contracts on what has now become the Chicago Mercantile Exchange. It was not until the early 1970s that exchanges began trading derivatives on financial assets, specifically foreign currency futures on the International Monetary Market and stock options on the Chicago Board Options Exchange. Within a decade, major exchanges added derivative products on stock indices and interest rate instruments. While new financial derivative products continue to appear, the start of the twenty-first century finds exchanges developing contracts on the two remaining major asset categories, human capital and real estate.[1]

Consider first the importance of managing the uncertain cash flows derived from human capital. In terms of relative size, annual gross domestic product (GDP) is approximately 10 times larger than corporate earnings. Moreover, Bottazzi, Pesenti, and van Wincoop (1996) find low correlation between returns on financial and human capital, so that existing financial derivatives would provide little in the way of managing risks associated with aggregate income. Citing the necessity for new macroeconomic contracts, Shiller (2003, p. 1) states: "We need to extend the domain of finance beyond that of physical capital to human capital, and to cover the risks that really matter in our lives." He goes on to argue that the trading of macroeconomic risks would allow individuals to pursue careers of their choice without fear of financial ruin and thus lead to societal gains and increased economic efficiency.

In addition to derivatives on income aggregates, markets might trade indexes based on a number of related macroeconomic variables. Examples might include (but are not limited to) contracts on inflation or production. Thus, we wish to consider more generally products on a number of macroeconomic indexes in the discussion of emerging derivative instruments. While macroeconomic indexes form the basis for these instruments, hereafter we refer to them simply as economic derivatives.

Real estate is the other large, outstanding asset class with a need for effective risk management instruments. At the end of 2005, the value of homes in the United States equaled \$21.6 trillion, compared to the \$17.0 trillion held in domestic equities and \$25.3 in fixed income securities (Labuszewski, 2006). Like financial assets, ownership of residential real estate exposes the investor to substantial price risk. Anecdotally, this can be seen by looking at the recent experience in Las Vegas house prices. As measured by the S&P/Case-Shiller® index, home prices dramatically increased 60.9 percent between January 2004 and August 2006 before reaching a peak and falling 41.2 percent over the next 29 months. Similar stories of bursting price bubbles appear in other regions of the country and underscore the need for effective risk management tools in real estate.

The remainder of this chapter discusses recent examples of economic derivatives, their trading experiences, and the potential for future success. Many of these derivatives are contingent claims that have a positive payout if an economic event occurs and traditionally have traded in parimutuel markets. The specifics about contract design and pricing of economic derivatives is left for Chapter 12. Instead, the next section considers market participants who might use economic derivatives and focuses on their ability to either hedge or speculate in these markets.

A similar discussion follows for potential users of real estate derivative products. This section also describes a popular index that forms the basis of several real estate derivatives. A final section concludes with remarks about the future of macroeconomic and real estate derivatives and the prospect of product innovation in these markets.

ECONOMIC DERIVATIVES

The first example of an exchange-traded economic derivative was the Consumer Price Index (CPI) futures launched in 1985 by the Coffee, Sugar and Cocoa Exchange (CSCE) in New York. While much fanfare in the business press preceded its launch, the inflation futures contract generated little trading interest and eventually died. Horrigan (1987) cites three factors for the low volume: (1) the relevancy of the CPI as a measure of inflation, (2) other preferable cross-hedges for inflation, and (3) no underlying asset for the futures contract. However, Horrigan thought a fourth reason, relatively little inflation risk during this period, was the main factor for the lack of trading activity.

Todd Petzel, the CSCE's chief economist at the time of the CPI futures launch, believes that liquidity, or the lack thereof, ultimately led to the contract's demise. Petzel (2001) observes that hedging and speculation are equally important in the ultimate success of a futures contract. Hedging by itself is not sufficient to ensure success; instead, speculators must provide liquidity to take the other side of a market where hedgers are net short or net long. Economic wisdom suggests that speculators will supply capital instantly if they perceive a mispricing in the market. However, as Petzel points out, the problem with this thinking is that it is not easy to tell when markets are mispriced so that speculators must have some exit strategy to minimize their losses.

For dealer-brokers who might supply speculative capital, one way to minimize these risks is to enter spread positions *across* markets. By monitoring the basis between related markets, traders are more likely to present bids and offers and supply the necessary liquidity to a new market without incurring too much risk. Petzel believes that the lack of a spread vehicle in the 1980s ultimately led to the failure of the CPI futures contract on the CSCE.

The growing supply of inflation indexed bonds over the last two decades is perhaps one factor that led the Chicago Mercantile Exchange (CME) to list a new CPI futures contract in 2004. The launch followed a Bond Market Association survey (2003) of dealers, financial institutions, and money managers that found more than 70 percent of respondents would utilize a CPI futures contract. However, the ability to form a three-legged arbitrage (indexed bonds, traditional nominal debt, and CPI futures) and the encouraging survey results were not enough to ensure the contract's popularity, and it ceased trading in 2006. The reason for this latest failure appears to have come full circle as the CPI futures contract (again) came out at a time when there were few concerns about hedging inflation.

The CPI futures contract was not the CME's only foray into economic derivatives. Following inflation futures trading, the CME teamed up with Goldman Sachs in 2005 to list derivative contracts on a number of different macroeconomic indexes. This union built on the over-the-counter market put together by Goldman Sachs and Deutsche Bank in 2002 and utilized clearinghouse procedures that minimized counterparty risk. Economic contracts traded on core inflation, U.S. gross domestic product, initial jobless claims, the Institute for Supply Management (ISM) manufacturing index, international trade balance, retail sales (excluding autos), non-farm payroll, and the euro area harmonized index of consumer prices. The exchange listed contracts for each index release date and used a Dutch auction process prior to the announcement.[2] The most popular index, non-farm payroll, averaged only $9 million per contract (Gadanecz, Moessner, and Upper 2007), and the anemic volume led the CME to cease trading of economic derivatives in June 2007. However, interest in economic derivatives continues, and in the spring of 2008, the CME again launched non-farm payroll futures and option contracts.

At a minimum, for these contracts to be successful, they must attract both hedgers and speculators. For the former group, the contracts must provide an effective hedge for their risks; for the latter group, there must be other instruments to spread the risk. In both cases, the returns to the economic derivatives need to be highly correlated with a relevant set of cash flows.

Consider correlation within the context of a state wishing to hedge its tax revenues to ensure full funding of government programs. For example, Alabama dedicates income and sales tax revenue to the Education Trust Fund (ETF). The fund finances all public education in the state, and theoretically, any economic derivative correlated with Alabama personal income should provide an effective hedge of tax revenues.

Since U.S. non-farm payroll was the most popular of the economic derivative contracts, suppose that Alabama uses it to hedge state income. Exhibit 16.1 shows the relation between the two time series from 1990 to 2006. Whereas the quarterly change in U.S. non-farm payroll is a relatively smooth time series that exhibits positive, first-order autocorrelation, Alabama personal income is much more volatile. The correlation between the two series is low (equal to .1794), and the cross-hedge

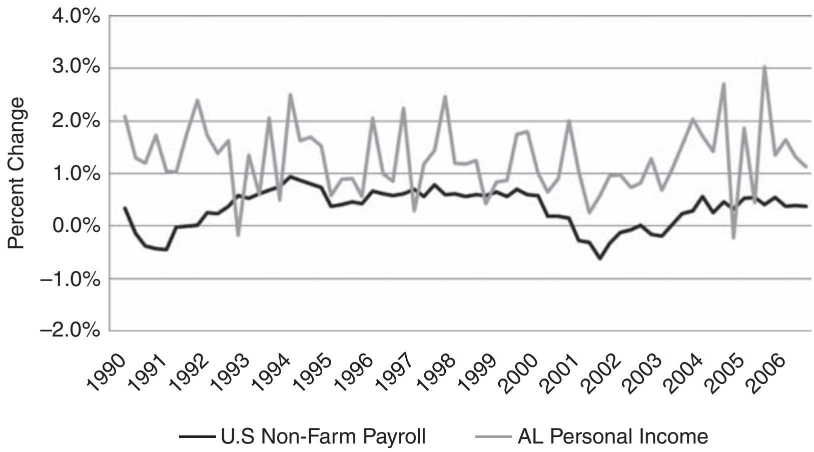


Exhibit 16.1 Quarterly Change U.S. Non-Farm Payroll and Alabama Personal Income

basis risk between US non-farm payroll and Alabama personal income is too large to provide an effective hedge of state revenue.

If the economic risks of the potential hedger are largely idiosyncratic and do not closely correlate with an economic indicator, index derivatives will provide little in the way of risk management. Without hedgers, an economic derivatives market will fail. The issue of basis risk again appears when looking at the market for real estate derivatives.

REAL ESTATE DERIVATIVES

There have been a number of attempts to trade real estate securities. Stocks and bonds traded on the New York Real Estate Securities Exchange (NYRESE) beginning in 1929, but the Securities and Exchange Commission decertified the NYRESE in 1941 following the plunge in real estate prices. In the early 1990s, residential and commercial real estate futures contracts briefly traded on the London FOX Property Futures Market. More recently, the Chicago Mercantile Exchange began trading options and futures contracts on residential housing indexes for 10 different cities and on a composite index. Subsequently, the CME started trading commercial real estate derivatives on five geographic regional indexes, a composite index, and four national property type indexes (office, warehouse, apartment, and retail).

Despite nearly two years of listing, CME residential real estate futures have met with limited success. In one recent day of trading, the composite index futures had 0 volume and open interest of 64 contracts. On the same day, the Las Vegas futures also had 0 volume and open interest of 49 contracts, this despite the city being one of the more dynamic housing markets in the United States over the last five years.

One reason for the lack of activity can again be traced back to basis risk and hedging effectiveness. Here the cross-hedge basis risk depends on the difference between the hedger's real estate portfolio returns and the returns to the CME real estate futures contract. The underlying index for the futures contract is the

S&P/Case-Shiller Metro Area Home Price Index. One important characteristic is that it is a repeat sales index so that a house enters the index only after it has been sold twice.

To see whether the futures contract might provide an effective hedge, Bertus, Hollans, and Swidler (2008) consider the risk management problem for three different groups of investors. The first group includes real estate investors holding equity stakes in property for returns generated by income and appreciation and mortgage portfolio investors looking to hedge mortgage default rates. A second group includes individual homeowners worried about the equity value of their home, while the third group contains real estate developers/builders who might hedge against price drops while holding an inventory of houses at various stages of completion.

Given the significant price swings in the Las Vegas real estate market this decade, the analysis examines sales in the metropolitan area and the ability of each investor group to manage their risk. Exhibit 16.2 illustrates hedging effectiveness for the three investor groups, where hedging effectiveness equals the percentage of price risk reduced by entering into a futures hedge using the CME contracts. The hedge period is a quarter of a year, and the graph depicts hedging effectiveness over a five-year (20-quarter) period.

The professional real estate investors (group 1) hold a diversified portfolio of properties in the Las Vegas area. The relevant return comparison is between the typical (median) house in Las Vegas and the return to the CME futures contract. From Exhibit 16.2, the CME futures contract would have reduced risk only by 30 percent or less during the latter half of the 1990s. Only in the last five-year period would

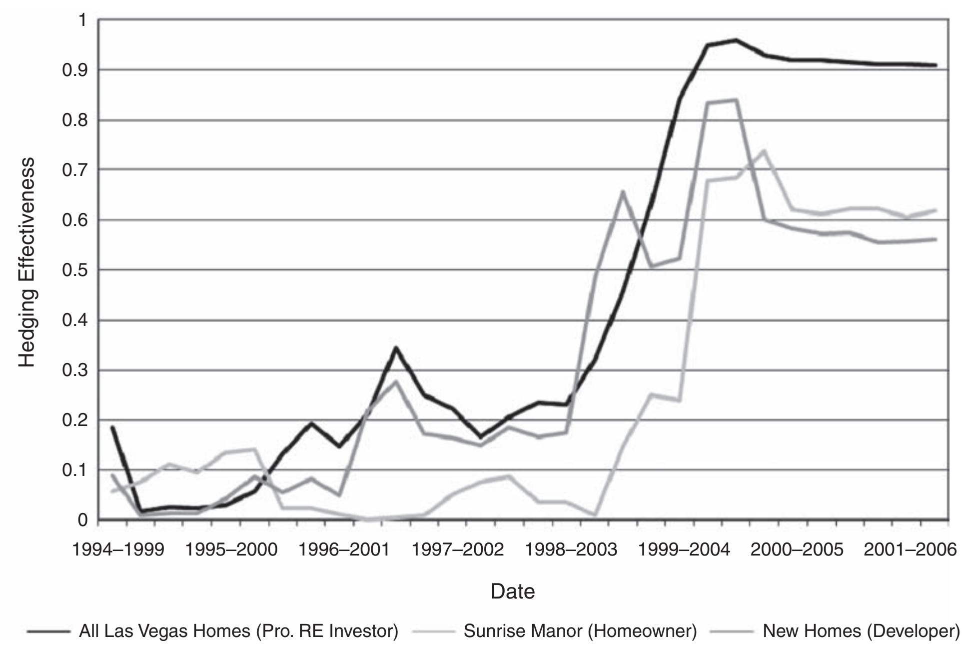


Exhibit 16.2 Hedging Effectiveness

CME futures have provided an effective hedge and eliminated approximately 90 percent of price risk for professional real estate investors.

Next consider a homeowner in Sunrise Manor, a low-income, unincorporated area within metropolitan Las Vegas. Correlation between home prices in this district and the CME futures contract are low, and even in the last five years, a homeowner would have reduced risk only by 60 percent. The implication is that house prices do not move uniformly across a metropolitan area, and individual homeowners may find that CME futures contracts do not provide an effective vehicle to manage idiosyncratic risk. Factors such as age and quality induce additional idiosyncratic risk and further exacerbate the risk management problem.

Finally, Exhibit 16.2 shows that for group 3, developers building new subdivisions, the CME futures contracts offer little in the way of hedging effectiveness. Even in the latter years of the sample, barely 50 percent of new-house price volatility in Las Vegas can be reduced with CME futures. Again, cross-hedge basis risk is substantial and is the result of new house prices not being in the S&P/Case-Shiller repeat sales index.

For many real estate investors, the cross-basis risk between their portfolio and a traded real estate index is large and suggests that they are unlikely to use the derivative contracts with any regularity. Moreover, arbitrageurs have limited opportunities as well given the illiquid nature of the cash market. Similar to economic derivatives, the question that arises is: What does the future hold for these securities?

THE NEXT FRONTIER

The slow acceptance of derivative products in a new market is nothing new. Stock index futures got off to a slow beginning when they first started trading on the Kansas City Board of Trade (KCBOT) in 1982. The KCBOT's Value Line contract was based on a geometric index of approximately 1,650 stocks that made it difficult to price. The Standard & Poor's 500 stock index futures that subsequently starting trading on the CME has become one of the world's most popular derivatives, but not before arbitrageurs found they could replicate the contract with a stock portfolio consisting of significantly fewer than 500 companies. Even the popular Treasury and Eurodollar futures have their roots in the now-defunct *G*overnment *N*ational *M*ortgage *A*ssociation (GNMA) contracts, the first exchange-traded derivative on interest rate securities.

While contract redesign and analytical solutions leading to lower-cost hedging and arbitrage opportunities may eventually lead to successful exchange-traded economic and real estate derivatives, the next round of activity in these areas is likely to include a variety of structured products. Already, a number of securities linked to GDP and real estate indexes exist and trade either in the over-the-counter market or with the help of a financial intermediary. These products are often tailor-made to address the specific risks of the issuer and thus mitigate the problem of cross-basis risk.

Dollar-denominated Brady bonds sold on behalf of Costa Rica in 1990 is one of the first examples of a GDP-linked security (Filippov 2005). On certain issues, interest payments would be increased if real GDP exceeded 120 percent of the country's 1989 level. In fact, Costa Rica real GDP passed the income threshold in 1993, and subsequently, the country decided to retire the index-linked bonds early.

In 1994, Citibank helped Bulgaria issue a nearly $1.9 billion discount bond with the feature of Additional Interest Payments (AIPs). The payments would occur if Bulgaria's GDP exceeded its 1993 level by more than 125 percent and there was a year-over-year increase in GDP. As it turns out, Bulgaria's economy remained weak throughout the decade, and the country never paid any additional interest. Having implicitly sold bondholders a call option on Bulgarian income that eventually expired worthless, the country saved substantially on its borrowing costs.

An example of an index-linked corporate security is Swiss Re and Societe Generale's underwriting of Compagnie Financiere Michelin bonds in 2000. Wishing to hedge its revenue, which is highly correlated with the income in its principal markets, Michelin issued debt tied to the GDP growth rates of the United States and several European countries. In effect, these bonds served as insurance for the company and allowed for restructuring of the debt if average GDP did not grow at a sufficient rate.

The foregoing examples largely solve the basis risk in a cross hedge as the hedging benefits of the debt are directly tied to the issuer's risk. Nevertheless, these products have inherent problems of their own, especially with respect to contract design. In particular, the relevant index used to trigger payments or restructuring must be clearly defined. In the case of Bulgaria's debt, for example, the index values were to be taken from the "World Tables of the World Bank," but it was not clear whether the relevant index was in constant or current units nor whether it was denominated in U.S. dollars or local currency.

Even if the index is carefully delineated in the bond payment terms, frequently there is an issue surrounding revisions of the index. Macroeconomic measures such as GDP often go through several revisions, each one potentially changing the payoff of the security. To address this issue, the payment scheme must make clear whether security returns are a function of the first release of the index or some later revision. With respect to sovereign debt, there is the related problem of index manipulation. If, for example, additional payouts occur if a threshold is crossed, the federal agency that calculates the index may feel pressure to manipulate the numbers and mitigate borrowing costs. Moreover, government policies may lead to economic outcomes that also affect index values, and thus, security payouts are subject to problems of moral hazard.

Moving on to real estate securities, we have seen that cross-basis risk is a particularly difficult problem to address in the risk management process. Even when drilling down to a metropolitan region, there is frequently low correlation between returns on a single house or portfolio of homes in the area with the returns of the metropolitan index. Perhaps acknowledging property idiosyncrasies within a real estate market, S&P/Case-Shiller have begun publishing indexes by price tiers. Thus, while the CME futures contracts trade only on city indexes, an investor could write over-the-counter contracts based on the low-tier, middle-tier, or high-tier price index for the metropolitan region.

While segmenting a region's index by price tiers mitigates cross-basis risk for some investors, house price dynamics are also influenced by a property's age, size, quality of material, and other attributes. One way to address these differences is through some sort of hedonic modeling.[3] Then the returns to a house or portfolio of homes can be compared to expected price appreciation, where expected price is determined by substituting the asset's attributes into a hedonic model.

Zurich Cantonal Bank issues index-linked mortgages and is one example of a note that may be used to hedge real estate risk (Syz, Vanini, and Salvi 2008). At the end of five years, the bank substitutes the attributes of the home into a hedonic model to estimate the value of the home and then compares this value to the loan's original balance. If the house value has declined, the homeowner is insured for the difference, and the balance on the new loan is adjusted to reflect the loss. The homeowner pays the premium for the embedded put option in the mortgage by accepting an interest rate slightly higher than the rate on a standard five-year note.

The example of the index-linked mortgage involves a financial intermediary in the risk management process. With real estate, use of an intermediary may evolve as the standard practice. Investors may not have the necessary expertise or may not hold a well-diversified portfolio to effectively hedge their real estate assets using standard derivative instruments like those listed on the CME. However, the bank issuing the index-linked mortgage does have the needed expertise. Moreover, the bank's portfolio of mortgages will be written on a diversified set of homes in an area so that the bank, in turn, can use products like CME derivatives to lay off its risks.

Derivatives on economic and real estate indexes are still in the very early stages of growth. To date, the trading activity in these markets has been modest, largely the result of significant cross-basis risk that affects both hedgers and arbitrageurs. However, the need to manage these risks is too large for the markets to disappear. In time, new instruments will develop and expertise will evolve that eventually will ensure the success of these markets.

ENDNOTES

1. Exchanges and over-the-counter markets continue to innovate in the financial derivatives area. Credit derivatives are one example of an emerging financial market. More information can be found in Chapter 13.
2. The exchange used the Longitude Auction Platform® to trade these economic derivatives. Longitude® technology is now owned by the International Stock Exchange (ISE), which announced plans to utilize this format for contracts on the ReXX Commercial Property Index in March 2008. In this parimutuel market, there is no need to match buy and sell orders as prices are determined by relative demand for related derivatives. Subsequent to its March 2008 announcement, the ISE determined that the alternative markets area, including the ReXX contracts, would not provide an adequate return on investment and turned its interest to other products with greater revenue potential.
3. A hedonic model is a way of estimating price or value based on the constituent components of the asset. For housing, price might be a function of size, age, bedrooms, bathrooms, and location. For a further discussion of hedonic models in real estate, see Malpezzi (2002).

REFERENCES

Bertus, M., H. Hollans, and S. Swidler. 2008. "Hedging House Price Risk with CME Futures Contracts: The Case of Las Vegas Residential Real Estate," *Journal of Real Estate Finance and Economics* 37, no. 3: 265–279.

Bond Market Association. 2003. TIPS survey. http://archives1.sifma.org/research/tips_survey.pdf.

Bottazzi, L., P. Pesenti, and E. van Wincoop. 1996. "Wages, Profits and the International Portfolio Puzzle," *European Economic Review* 40, no. 2: 219–254.

Chance, D. 1998. *Essays in Derivatives*. New York: John Wiley & Sons.

Filippov, A. 2005. *Macroeconomic Derivatives: Overview and Sovereign Debt Market Applications*. Master's thesis, Norges Handelshoyskole, Spring.

Gadanecz, B., R. Moessner, and C. Upper. 2007. "Economic Derivatives," *BIS Quarterly Review* (March): 69–81.

Horrigan, B. R. 1987. "The CPI Futures Market: The Inflation Hedge that Won't Grow," *Business Review—Federal Reserve Bank of Philadelphia* (May/June): 3–14.

Labuszewski, J. 2006. "Introduction to CME Housing Futures and Options," Chicago Mercantile Exchange Strategy Paper, 1–28.

Malpezzi, S. 2002. "Hedonic Pricing Models: A Selective and Applied Review," in K. Gibb and A. O'Sullivan, eds., *Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*. London: Blackwell.

Petzel, T. 2001. "Elusive liquidity," *@Markets Magazine* (January/February).

Shiller, R. 2003. The *New Financial Order: Risk in the 21st Century*. Princeton, NJ: Princeton University Press.

Syz, J., P. Vanini, and M. Salvi. 2008. "Property Derivatives and Index-Linked Mortgages," *Journal of Finance, Real Estate and Economics* 36, no. 1: 23–35.

SUGGESTED FURTHER READING

Additional historical examples of forward and futures markets can be found in:

Culp, C. 2004. *Risk Transfer, Derivatives in Theory and Practice*. Hoboken, NJ: John Wiley & Sons.

Readers interested in contract innovation and the problems of designing exchange-traded contracts in new markets can check:

Johnston, E., and J. McConnell. 1989. "Requiem for a Market: An Analysis of the Rise and Fall of a Financial Futures Contract," *Review of Financial Studies* 2, no. 1: 1–23.

Penick, M. 2004. *The Life Cycle of Futures Contracts: The Success and Failure Rates of Futures Contracts in the United States*. Working paper, CFTC, Washington, DC.

Thomas, S. 2002. "The Saga of the First Stock Index Futures Contract: Benchmarks, Models and Learning," *Journal of Money, Credit and Banking* 34, no. 3: 767–808.

An article discussing the practical issues of government hedging is

Hinkelmann, C., and S. Swidler. 2005. "State Government Hedging Using Financial Derivatives," *State and Local Government Review* 37, no. 2: 127–141.

A case for developing countries using GDP-linked bonds to hedge is made in:

Griffith-Jones, S., and K. Sharma. 2006. *GDP-Linked Bonds: Making It Happen*. United Nations, Department of Economic and Social Affairs working paper no. 21 (April).

ABOUT THE AUTHOR

Steve Swidler is the Mackin Professor of Finance at Auburn University. Prior to joining the Auburn faculty, he taught at the University of Texas at Arlington, Southern Methodist University, the University of Wisconsin-Milwaukee, and Rice University. Professor Swidler has also had international appointments at Victoria University in New Zealand, the Oslo School of Business, the University of Vaasa in Finland, and Tilburg University in the Netherlands. In addition to his academic experience, he has worked at the Office of the Comptroller of the Currency and at Lexecon, an economic consulting group. Professor Swidler obtained his undergraduate degree from Oberlin College and received his PhD in economics from Brown University. His teaching and research interests include financial engineering, investments, and security analysis, and he has published in a number of professional journals, including the *Journal of Finance;* the *Journal of Money, Credit and Banking*; and the *Journal of Financial and Quantitative Analysis.*

PART III

The Structure of Derivatives Markets and Institutions

Mastery of any market requires an understanding of the operations of the market, including its participants, its regulation, and the procedures and risks inherent in it. Part Three presents a detailed overview of these factors for derivatives markets, particularly as they exist in the United States.

Michael A. Penick analyzes current conditions in Chapter 17, "The Development and Current State of Derivatives Markets." After discussing the development of futures markets in the United States during the nineteenth century, Penick quickly moves to the current period. He covers international markets, which have developed very quickly over the last decades to challenge U.S. markets. Two of the most important developments that he covers are the rise of electronic trading and the amazing growth of the over-the-counter market.

In financial markets as they exist today in the United States, including derivatives markets, numerous market intermediaries functioning to connect the ultimate buyer and seller. These intermediaries are necessary to the function of the market, as James L. Carley explains in Chapter 18, "Derivatives Markets Intermediaries: Brokers, Dealers, Pools, and Funds." As he notes, these intermediaries fall into two broad categories: those who provide transaction execution services and those who provide money management services. Carley explains the different types of intermediaries, the functions each performs, and the regulatory environment in which they operate.

In Chapter 19, "Clearing and Settlement," James T. Moser and David Reiffen explain the nature of these processes for the trading of derivatives on organized exchanges. In these markets, a clearinghouse plays a critical role in clearing and settling trades. In simplest terms, once a trade is executed on the exchange, the clearinghouse substitutes its creditworthiness for that of each of the traders and guarantees performance to both traders. In the process of explaining how clearinghouses function, Moser and Reiffen also explain the system of margin that is such a dominant feature of trading derivatives on organized exchanges.

One of the serious problems in the financial crisis that began in 2007 is that of trust between buyers and sellers of financial instruments, which translates into problems of counterparty credit risk—the risk that your opposite trading party will default on his or her obligations. Previously a topic of interest to only a few market specialists, the problem of counterparty credit risk has come to the fore during the recent crisis, as James A. Overdahl explains in Chapter 20, "Counterparty Credit Risk." Counterparty credit risk has always been an issue, but many participants in the market had impeccable credit. However, during the recent credit crisis, the credit quality of AAA-rated firms so quickly fell that counterparty credit risk became an all-consuming concern. Overdahl traces the issues involved in measuring counterparty credit risk and discusses the ways in which this kind of risk affects the behavior of participants in derivatives markets.

All through the credit crisis, the organized commodity futures and option exchanges in the United States have performed more or less flawlessly with no significant defaults. These markets are, in fact, closely regulated, as Walter L. Lukken explores in Chapter 21, "The Regulation of U.S. Commodity Futures and Options." As he explains, the dominant regulatory system for futures and futures options was set by the Commodity Exchange Act, which dates back to 1922. In 1974, Congress established the Commodity Futures Trading Commissions, which is charged with regulating these markets. Lukken explains the scope and process of these laws and their accompanying regulations.

As the reader has surely gleaned, financial derivatives have their own peculiarities, so it is not surprising that accounting rules for financial derivatives are somewhat specialized, as Ira G. Kawaller explains in Chapter 22, "Accounting for Financial Derivatives." The overarching accounting principles for financial derivatives are articulated in Financial Accounting Standard No. 133. Kawaller explains how these standards are applied to various transactions and discusses the specialized accounting rules that apply to qualifying hedging transactions.

In spite of their obvious and important value and their contribution to the economy, financial derivatives come to the public's attention most prominently when things go wrong. And there certainly are periodic spectacular disasters, as John E. Marthinsen chronicles in Chapter 23, "Derivatives Scandals and Disasters." Marthinsen analyzes five of the most sensational derivatives mishaps of recent years, not merely to tell a highly entertaining story but also to identify similarities among these debacles and to draw lessons for strengthening these markets. The five derivatives scandals range from 1993 to 2008, culminating with the loss of $7.2 billion in 2008, which was caused by a single young trader at the French bank Société Générale. In all, losses in just these five events totaled more than $20 billion.

CHAPTER 17

The Development and Current State of Derivatives Markets

MICHAEL A. PENICK
Senior Economist, Commodity Futures Trading Commission

INTRODUCTION: THE SITUATION IN THE 1960s

In describing the growth of derivatives trading, it is useful to have a point of reference. This chapter begins with the 1960s as this decade bridged the "old" world of commodity-based derivatives trading that existed for nearly a century and the "new" world of financial derivatives that emerged in the 1970s. As the 1960s dawned, the futures industry in the United States looked much the same as it had at the end of the nineteenth century. According to the volume reports of the Association of Commodity Exchange Firms (predecessor to the Futures Industry Association [FIA]), in 1960, there were 14 futures exchanges in the United States trading a total of about 3.9 million contracts. The Chicago Board of Trade (CBOT) was, as it had always been, by far the largest exchange with a volume of 2.5 million contracts (of which 1.2 million were in the largest contract, soybeans). The CBOT also traded significant volumes of wheat, corn, oats, rye, soybean oil, and soybean meal. The Chicago Mercantile Exchange (CME) was the second largest U.S. exchange with a volume of 567,000 contracts, almost all in eggs. The third largest exchange was the New York Mercantile Exchange (NYMEX), almost all in potato futures. Other exchanges included the Chicago Open Board of Trade (later the MidAmerica Commodity Exchange), which traded miniature versions of CBOT products; the Kansas City Board of Trade (KCBT); the Milwaukee and Minneapolis Grain Exchanges; the New Orleans and New York Cotton Exchanges; the New York Produce Exchange; the Memphis Board of Trade, the New York Cocoa Exchange; and the New York Coffee and Sugar Exchange. Agricultural commodities accounted for 97.5 percent of U.S. futures trading, but the Commodity Exchange in New York traded various metals and industrial products.

Futures trading began in the United States following the founding of the CBOT in 1848. The CBOT was initially a cash grain market, but over time, the elements

I have benefited from the comments of Bob Kolb, Jim Moser, Jim Overdahl, and Rick Shilts. The views expressed in this chapter are those of the author and should not be viewed as those of the Commodity Futures Trading Commission, its commissioners, or its staff.

of a modern futures contract were introduced.[1] The CBOT considered the first true futures contract to have been traded in 1877. During the nineteenth century, exchanges trading grain futures appeared in many cities. The CME and NYMEX were founded to trade butter and eggs. By 1900, most of the exchanges that existed in 1960 had been founded. Moreover, most of the commodities that were being traded in 1960 dated back to the nineteenth century.[2] Futures exchanges in the United States conducted their trades through the open outcry system, which also dated back to the nineteenth century. Open outcry is a method of public auction where trading occurs in a pit or ring on a trading floor and traders may bid and offer simultaneously either for their own accounts or for the accounts of customers.

The futures industry in the United States seemed stagnant in 1960 and was unimportant in the rest of the world. Volume had declined since the FIA's predecessor began compiling data in 1954. While volume data from the 1920s and 1930s is incomplete, it appears that in 1960 the futures industry was smaller in terms of both volume and the number of exchanges than it had been between the world wars.[3] The 1958 ban on onion futures had not helped; it left the second and third largest exchanges, the CME and NYMEX, with only one commodity each. Nevertheless, the industry was about to embark on nearly a half century of incredible growth and innovation. By 2007, the CME's Eurodollar futures and options contracts alone were trading an average of 3.22 million contracts per day—almost as much as the entire futures industry annual volume in 1960. In 2007 and 2008, many billions of futures and options were traded on exchanges worldwide in a variety of agricultural, energy, metal, and especially financial commodities. The off-exchange or over-the-counter (OTC) derivatives markets were even larger than the exchange-traded markets. The derivatives markets, for good or ill, played a major role in the world economy of the early twenty-first century. This chapter describes how the industry evolved from the sleepy agricultural hedging markets of the 1960s and earlier to the behemoth of the early twenty-first century.

The primary innovation of the 1960s was the introduction of livestock futures by the CME, beginning with pork bellies in 1961 and followed by live cattle in 1965 and live hogs in 1966. Pork bellies exhibited low volume at first, but in 1969, the trading volume of pork belly futures exceeded 2 million contracts, more than any other futures contract and comprising almost a quarter of the futures industry's volume that year. The live cattle and live hog contracts were also successful. With egg futures markets in sharp decline for much of the 1960s, it is questionable whether the CME would have survived without the success of the livestock complex. New kinds of agricultural futures had been introduced before (the most successful was soybeans in 1936), and there were other (unsuccessful) attempts at new futures contracts during the 1960s, including mercury and apples. The introduction of livestock futures does not seem groundbreaking in and of itself. However, by ensuring the CME's survival at a critical time, the livestock complex demonstrated the value to exchanges of innovation and paved the way for the much greater innovations of the 1970s and 1980s.

FINANCIAL FUTURES AND OPTIONS

The introduction of financial futures during the 1970s was the most important innovation in the history of futures trading. Volume grew explosively and financial futures dominated the industry by the 1980s. The first financial futures were on foreign currencies. These were followed by equity options, interest rate futures, and eventually stock index futures and options.

It is generally believed today that the CME invented foreign currency futures. While the CME was the first exchange to succeed with currency futures, the first exchange to list them was the International Commercial Exchange. The International Commercial Exchange is nearly forgotten today,[4] but the FIA's predecessor's volume reports indicate that they began trading futures contracts on major currencies such as the British pound and Japanese yen on April 23, 1970, two years before the CME. Volume was very low. The Bretton Woods system of fixed exchange rates was still in effect in 1970, and there was not yet a need for currency futures. By May 16, 1972, when the CME launched its complex of currency futures with support from economists at the University of Chicago,[5] the Bretton Woods system was collapsing, and the CME met with more success. The International Commercial Exchange disappeared in 1973. Other exchanges also listed currency futures, but the CME consistently has had the largest market share in currency futures. The CME's market share in the overall currency market always has been small compared to the overall OTC interbank market. Although the currency futures market has grown substantially since the advent of electronic trading (discussed later in this chapter), most large trades still are done in the interbank market. The currency futures market historically has been a niche market for small hedgers and retail speculators. However, the currency market is very large. Even though the futures market share has been small, the volume has been large enough for these contracts to be successful.

While currency futures were important, it was the next innovation, interest rate futures that enabled financial futures to transform the futures industry. The volume in interest rate futures soon dwarfed that of the traditional agricultural commodities. In the past, futures markets were separate from securities markets, and futures commission merchants typically were not the same firms as the stock brokers and investment banks that participated in the securities industry. Now those securities firms became major participants in the futures markets as well.

The first interest rate future was on Ginnie Mae CDRs (collateralized depository receipts), launched by the CBOT in October 1975. Ginnie Mae CDRs were mortgage-backed certificates that were guaranteed by the Government National Mortgage Association (Ginnie Mae). In January 1976, the CME introduced futures on three-month Treasury bills. While these contracts initially were moderately successful, both are gone today. It was the CBOT's long-term (15–30 year) bond futures contract and the three-month Eurodollar futures contract that truly altered the futures landscape. The long-term bond futures contract was launched in August 1977 and was the largest contract on the CBOT by 1981. It was followed by contracts on short-term (2 years), medium-term (5 years), and long-term (7–10 years) Treasury notes. Today, the 5- and 10-year note futures are more popular than the 30-year bond futures,[6] but the Treasury complex remains one of the highest-volume groups of contracts.

The three-month Eurodollar contract was the first cash-settled futures contract (currency and Treasury futures typically contracts provide for delivery via wire transfer) and was launched in December 1981. It is based on the three-month London Interbank Offered Rate (LIBOR), which is calculated each day by the British Bankers Association. During the 1990s, the Eurodollar contract surpassed the

Treasury bond contract to become the largest futures contract in the United States and remains so today. As will be discussed, the Eurodollar contract has proven to be a good hedging mechanism for interest rate swaps. The cash settlement feature has proved to be an important innovation. It is commonly used where physical delivery is impracticable or impossible. Some former physical delivery contracts (such as pork bellies) have converted to cash settlement because of limitations on deliverable supply.

Meanwhile, the Chicago Board Options Exchange (CBOE) was founded in 1973 by members of the Chicago Board of Trade to trade options on individual equities under the jurisdiction of the Securities and Exchange Commission (SEC), but in an open outcry environment similar to the futures markets.[7] That same year, the Black-Scholes option pricing model was published in the *Journal of Political Economy.* The CBOE was highly successful, and stock exchanges such as the American Stock Exchange and the Pacific Stock Exchange established option markets using, like the stock exchanges, a specialist system.[8] Options on futures had been illegal since 1936, but options on financial futures were launched in 1982, initially as a pilot program. Options on agricultural futures were added in 1986. Soon most futures contracts had associated option contracts as a matter of course.

Cash settlement was essential for the last of the major types of financial futures—stock index futures. Regulatory conflict between the SEC and the futures market regulator, the Commodity Futures Trading Commission (CFTC), both of which claimed jurisdiction over equity futures, delayed the launch of stock index futures. While stock index futures were proposed in 1977,[9] it took until 1981 for the SEC and CFTC to come to an agreement. This agreement was known as the Shad-Johnson Accord, named after the agencies' respective chairmen. Under the accord, nonnarrow-based cash-settled stock index futures (and later, options on futures) could be traded on futures exchanges under CFTC jurisdiction while stock index options could be traded on securities exchanges under SEC jurisdiction. Futures on individual stocks were banned and remained illegal in the United States until 2000.[10] The first stock-index futures contract was on the Value Line Index, launched by the tiny KCBT in February 1982. The most successful contract by far, the Standard & Poor's 500 Stock Index (S&P 500) futures contract, was launched by the CME in April 1982. Dow Jones and Company refused for many years to permit futures on the Dow Jones Industrial Average (DJIA), but the CBOT launched a futures contract on the DJIA, with some success, in 1997.

Overall, by the mid-1980s, financial futures and options dominated the industry and dwarfed the volume in physical commodities, such as agricultural futures. Even the physical commodities were becoming dominated by a new product group: energy futures. The original futures contracts—agricultural futures—were becoming less and less important.

FOREIGN MARKETS

There have been futures contracts outside of the United States at least since the seventeenth century in Holland. There were also older commodity exchanges in Japan, Australia, and Canada, among other places. During the 1980s, new futures exchanges were founded to trade financial futures in most of the larger market economies in Western Europe and East Asia. Many of these exchanges exhibited growth rates exceeding those in the United States. Some existing stock exchanges, such as the Osaka Securities Exchange in Japan, began listing stock index futures contracts. Some of the more important foreign futures exchanges were: the London International Financial Futures Exchange (LIFFE, now part of NYSE Euronext), founded in 1982; the Deutsche Terminbourse in Germany (DTB, now part of Eurex), founded in 1988; the Tokyo International Financial Futures Exchange (TIFFE, now the Tokyo Financial Exchange), founded in 1989; and the Singapore International Monetary Exchange (SIMEX, now part of the Singapore Exchange), founded in 1984. Most of these exchanges listed futures contracts on their home countries' short-term interest rates (e.g,. three-month Euroyen at TIFFE), government bonds (e.g., the 10-year bund and gilt on LIFFE), and/or broad-based stock indexes (e.g., the DAX index at DTB and the Nikkei 225 at the Osaka Securities Exchange). Most of these contracts were highly successful.

For the most part, exchanges were not successful at listing products based on other countries' stock indexes or interest rates. For example, the CME's Nikkei 225 contract was always tiny compared to the Osaka contract, and LIFFE listed a U.S. Treasury bond contract during the 1980s that failed. There have been a few exceptions: LIFFE was the first exchange to list a futures contract on German government bonds and maintained a majority market share for several years after DTB listed a similar contract. Only when DTB went electronic in the late 1990s was it able to wrest the contract from LIFFE. SIMEX enjoyed some success with three-month Eurodollars for many years because it entered into a mutual offset agreement with the CME whereby market participants could enter into positions on one exchange and exit on the other. In contrast, the CBOT Treasury bond contract was not fungible with the LIFFE contract.

During the 1990s and 2000s, futures exchanges have been founded in former communist nations, such as Russia and Poland,[11] and in many other countries on every inhabited continent, including India, Brazil, and South Africa. Most of these exchanges specialize in local commodities and financial instruments. Growth worldwide has been enormous and has outpaced that of the United States. For several years prior to the 2007 merger of the CME and the CBOT, the largest futures exchange in the world was the Swiss-German Eurex.

OTC MARKETS

While futures markets exhibited dramatic growth worldwide during the 1980s and 1990s, they also faced significant competition from fast-growing off-exchange or OTC derivatives markets. The most popular OTC derivative is a swap. In general, a swap is an exchange of cash flows—for example, one currency for another (currency swap) or a fixed interest rate for a floating interest rate (interest rate swap). A swap is a privately negotiated agreement between two counterparties (commonly a swap dealer such as a bank and an end user) that typically provides for a series of specified payment dates over a specified time period. On each payment date, one counterparty pays the net difference between the cash flows to the other. The first currency swap was executed in 1980, and the first interest rate swap was executed in 1982. Major commercial banks and investment banks established large swap dealer operations. Currency and interest rate swaps became increasingly standardized. For example, a common (plain vanilla) interest rate swap might call for payments every three months for a three-year period. Economically, a swap (especially a vanilla swap) is equivalent to a futures contract or to a strip of futures contracts on short-term interest rates (such as the CME three-month Eurodollar contract),[12] but swaps are not traded on an exchange. The currency and the interest rate swap market grew to dwarf the exchange-traded futures market in notional value. According to the Bank of International Settlements (BIS), in June 2008, the outstanding notional principal in interest rate swaps was $357 trillion, while the outstanding notional principal in interest rate futures worldwide was $27 trillion.[13] On that date, the outstanding notional principal in currency swaps and forwards was $49 trillion, while the outstanding notional principal in currency futures was a relatively small $145 billion. There was also a substantial (about $10 trillion notional principal) OTC derivative market for equity and equity index swaps and options in June 2008. While BIS does not provide complete data for equity-related exchange-traded derivatives, the OTC equity derivative market is probably somewhat smaller than the exchange-traded market.[14]

In light of the economic equivalence between CME three-month Eurodollar futures contracts and certain interest rate swaps, the Eurodollar contract is used by banks and other swap dealers to hedge their residual positions. This has been a prime factor in making the Eurodollar contract by far the most actively traded futures contract in the United States in recent years. Similar short-term interest rate futures contracts in other countries (such as Japan's three-month TIBOR contract) also are actively used to hedge interest rate swaps. Thus, while the OTC derivative markets are larger than the exchange-traded futures markets, the exchange-traded markets have benefited in some ways from the existence of the OTC markets.

A more recent development in the OTC world has been the rise of the credit derivative markets. While there are several types of credit derivatives, the most common is the credit default swap (CDS), which essentially is an insurance policy on the debt issuance of a specific corporation, government entity, or other "reference entity." The seller of a CDS receives a payment from the buyer in exchange for agreeing to make a payment to the buyer in the event a specified credit event occurs on or before a specified date. A credit event might include a declaration of bankruptcy by the reference entity or a default on the reference entity's debt. The first credit derivatives were traded during the late 1990s. Credit derivatives grew rapidly during the first decade of the twenty-first century with outstanding notional principal nearly tripling from $20 trillion to $57 trillion just between June 2006 and June 2008, according to BIS. However, as will be discussed, credit derivatives played a major role in the financial crisis during the second half of 2008, and it appears likely that the credit derivatives market will shrink from the June 2008 level.

ENERGY DERIVATIVES

The most recent highly successful innovation in the exchange-traded futures markets was the introduction of energy futures contracts. There were futures contracts on petroleum as far back as the nineteenth century and again in the 1930s, but energy futures contracts became successful on a long-term basis only in the 1980s. The first modern attempt at an energy futures contract was a crude oil contract launched in 1974 (shortly after the 1973 Arab oil embargo and subsequent increase in oil prices) by the New York Cotton Exchange (a predecessor to ICE Futures US). This contract traded modest volume initially but was gone within two years. The New York Mercantile Exchange was more successful with heating oil in 1978 and leaded gasoline in 1981. Both contracts continue to trade actively to this day (with some modifications in the case of gasoline).[15]

During the early 1980s, both the CBOT and NYMEX worked to design a futures contract on crude oil that would be more successful than the New York Cotton Exchange contract. Both contracts received CFTC approval on March 29, 1983, and were launched the following day. It was the NYMEX contract that succeeded. This may be because NYMEX already had successful contracts on gasoline and heating oil and it was easier to trade crack spreads[16] on a single exchange, or it may be because the marketplace preferred NYMEX's West Texas Intermediate (WTI) contract over the CBOT's Louisiana-based contract. The NYMEX WTI crude oil contract became the largest futures contract on a physical commodity. Meanwhile, in London, the International Petroleum Exchange (IPE) was founded in 1980 with a gas oil (heating oil) futures contract and several other petroleum-based contracts. IPE's most successful futures contract, on Brent (North Sea) crude oil, began trading in 1988.

The last major exchange-traded energy futures contract in the United States was in natural gas, launched by NYMEX in 1990. The contract provided for delivery of natural gas at the Henry Hub in Louisiana. NYMEX launched several electricity futures contracts in 1996. These exhibited modest success for a few years but had disappeared by 2001. NYMEX tried again with electricity contracts in the 2000s, again with modest success.

During the 1980s and 1990s, OTC energy derivative markets also developed, especially in natural gas, where an enormous variety of contracts became popular. Many of these OTC derivatives were designed to be traded in conjunction with the NYMEX natural gas futures contract. For example, while the NYMEX contract provides for delivery only at the Henry Hub, natural gas prices at other U.S. locations often vary widely from prices at the Henry Hub. In light of this, basis swap contracts were developed that are cash settled on the difference between the NYMEX futures price on the last trading day and the spot price at a specified location such as Chicago or New York. A trader who needs to hedge the Chicago natural gas price can take a long position in both the NYMEX futures contract and the Chicago basis swap. This combination creates a synthetic futures contract on Chicago natural gas.

In 1999, the energy trading company Enron created an electronic trading platform called EnronOnline, which listed an enormous variety of energy derivatives as well as some nonenergy products, such as lumber and even broadband. The EnronOnline platform made trading easy and contributed to a substantial expansion of the energy derivatives market. Enron rival Dynegy created a similar platform called DynegyDirect, but it was not as successful. EnronOnline was a one-to-many platform whereby Enron was the counterparty to every participant rather than a traditional exchange. EnronOnline briefly dominated the energy derivatives market until Enron declared bankruptcy in late 2001.[17] Swiss bank UBS bought the EnronOnline platform but was unable to relaunch it successfully. DynegyDirect was shut down in 2002. In general, the OTC energy market seemed to be going into a tailspin in 2002. The creditworthiness of one's counterparty is important in the OTC markets, because OTC derivatives, unlike exchange-traded derivatives, traditionally are not cleared. Following the Enron bankruptcy, the creditworthiness of all of the OTC energy market participants was called into question and market volume declined sharply. There were two big beneficiaries to this event: a new exchange called the Intercontinental Exchange (ICE) and NYMEX.

ICE was founded in May 2000 by a consortium of energy trading firms and large banks. It began as a many-to-many electronic trading platform for OTC energy derivatives, typically the same products as were offered on EnronOnline. In 2001, ICE bought London's IPE and subsequently renamed that petroleum futures exchange ICE Futures Europe. When Enron declared bankruptcy, much of EnronOnline's business migrated to ICE. ICE benefited further as banks and investment banks entered the energy business and the OTC energy market recovered from the 2002 collapse.

NYMEX also benefited from the 2002 collapse in the OTC energy market, in part because some energy derivatives trading migrated from the OTC market to the exchange. As noted, futures contracts are cleared through a clearinghouse associated with an exchange, thus eliminating counterparty credit risk. The credit risk concerns in the OTC energy markets contributed to the volume increase at NYMEX. Moreover, in 2002, NYMEX established the ClearPort Internet-based platform on which market participants could either trade a wide variety of energy derivatives, or present privately negotiated energy derivatives transactions to ClearPort for clearing through the NYMEX clearinghouse. ClearPort proved successful although its market share in trading generally has been smaller than that of the ICE trading platform. ClearPort was the first major platform in the United States for clearing OTC derivatives. As discussed at the end of this chapter, the idea of OTC clearing has returned to the forefront recently as counterparty credit risk again became a major concern on a much larger scale in 2008.

THE RISE OF ELECTRONIC TRADING

As noted at the beginning of this chapter, futures exchanges in the United States traditionally have conducted their business in trading pits through open outcry auction. The foreign boards of trade that were launched in the 1980s also used the open outcry system. However, during the 1980s, as computers grew more powerful, exchanges began developing electronic trading systems, originally contemplated as supplements to the open outcry system. In 1992, the CME and CBOT jointly launched the GLOBEX electronic trading platform. The CBOT later left GLOBEX in favor of its own electronic trading platform but eventually returned to GLOBEX when the CBOT and CME merged. At first, electronic trading systems in the United States were used only for a supplemental trading session during the evening and overnight hours (enabling Asian and European investors to trade U.S. products during their trading days). The major exchanges in the United States were resistant to the idea of electronic trading during the U.S. trading day, because their members (who owned the exchanges) did not want to replace the open outcry system. The owner/members of futures exchanges had the privilege of standing in the pits and trading for their own accounts (such traders are known as locals) and/or filling orders on behalf of others (floor brokers). The locals benefited from the extra market information they obtained from standing in the pit (where traders can observe who is buying and selling and in what quantities), and floor brokers earned income from brokerage fees. In an electronic trading environment, these advantages would fade, being hard to replicate. In light of this, the members were able to delay the launch of electronic trading during the U.S. trading day at U.S. futures exchanges.

It is usually difficult for one exchange to compete with another exchange's established contract, once the established market has developed a high degree of volume and liquidity.[18] Most futures contracts have appeared to be natural monopolies with most volume and liquidity gravitating to a single exchange. The fact that futures contracts are not fungible (a futures contract entered into on one exchange cannot be offset on another exchange even if the terms are identical) makes it even more difficult to compete with an existing contract. There have been several failed attempts by various exchanges to list U.S. Treasury bond and/or note futures contracts to compete with the CBOT: LIFFE in the 1980s, Cantor Financial Futures Exchange in the 1990s, BrokerTec and Eurex US in the 2000s. While the power of incumbency remains strong, the advent of electronic trading appears to have improved the ability of exchanges to compete with incumbent exchanges.

Electronic trading was fully adopted overseas even as some U.S. exchanges resisted. In 1998, the Swiss-German exchange Eurex (the product of a merger between DTB and the Swiss Options and Financial Futures Exchange [SOFFEX]) scored the first big victory for electronic trading. For many years, the United Kingdom's LIFFE had dominated the market for German government bond (bund) futures. LIFFE listed bund futures in 1988. and the LIFFE contract was actively traded and well established when DTB launched its bund contract in November 1990. As is often the case, the exchange that listed first and established liquidity was able to maintain the lion's share of the activity in the contract. For most of the 1990s, the majority of the activity in the bund contract remained at LIFFE. It was the adoption of electronic trading by DTB successor Eurex that finally enabled it to wrest the bund futures contract from LIFFE. The bund is a major contract, and its loss was a big blow to LIFFE. Within a couple of years, LIFFE itself had become exclusively electronic and closed its trading floor. In the late 1990s and early 2000s, almost all derivatives exchanges outside the United States became exclusively electronic.

For reasons just discussed, U.S. exchanges were slower to adopt electronic trading. Two factors enabled electronic trading finally to dominate in the United States: demutualization and competition from electronic exchanges overseas. The three largest U.S. exchanges, the CME, the CBOT, and NYMEX, all converted during the early 2000s from member-owned mutual exchanges to publicly traded corporations where trading privileges were separated from ownership. This made it easier for exchange management to introduce side-by-side electronic trading, especially when overseas exchanges listed electronic versions of U.S. futures contracts.

The CME began adopting electronic trading earlier than other U.S. exchanges and started allowing side-by-side electronic and open outcry trading of Eurodollar futures during the U.S. trading day as early as 1999. Also during the late 1990s, the CME created "E-Mini" versions of its major stock index futures contracts, the S&P 500 and the Nasdaq 100. The E-Mini contracts have smaller contract sizes than the original contracts and were listed exclusively on the GLOBEX electronic platform. Today, the E-Mini contracts are significantly more active (adjusted for contract size) than the original contracts (which are available for trading both on the floor and electronically).

It was the need to fend off competition from Eurex US's electronic Treasury note contracts (which were listed by the U.S. subsidiary of Eurex in 2004) that caused the CBOT to allow side-by-side trading in its flagship Treasury contracts. Side-by-side trading in CBOT agricultural contracts began in 2006. NYMEX also began side-by-side trading of energy and metals futures in 2006, again because of electronic competition from ICE Futures Europe. The former IPE had listed a cash settled electronically traded WTI crude oil contract, which quickly captured about 40 percent of the market.[19] In 2007, ICE acquired the New York Board of Trade and announced plans to introduce electronic trading and close the futures pits. Once electronic trading was allowed at various exchanges, most futures trading migrated to the electronic platform, but options trading largely remained in the pits. While exchanges frequently offered enhancements to electronic trading systems for options, many traders still considered open outcry superior for executing complex option strategies. Overall, volume tended to increase substantially when electronic trading was introduced. As shown in Exhibit 17.1, the volume of futures and options traded on U.S. futures exchanges more than quintupled from about 600 million in 2000 to 3.2 billion in 2007. In contrast, volume increased by about 75 percent during the 1990s.[20] The CME's currency futures especially seemed to benefit from the switch to electronic trading, although the currency futures markets remained small relative to the OTC and interbank markets.

One effect of electronic trading is to make it possible for small "niche" futures contracts to survive on very low volume. A significant trading volume is needed to induce local traders and floor brokers to stand in a pit. A futures contract will die out if there are not enough trades. In an electronic environment, a futures contract can persist even if trades occur only occasionally. For example, the CME has listed

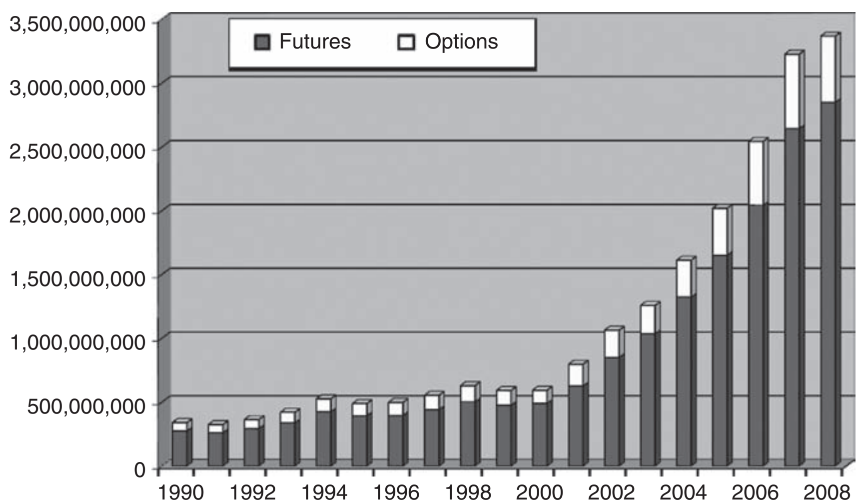


Exhibit 17.1 Total U.S. Futures and Futures Options Volume: 1990–2008
Source: Commodity Futures Trading Commission Division of Market Oversight.

on GLOBEX dozens of weather derivative contracts that are cash settled based on temperature or precipitation statistics in various U.S. or European cities. These contracts, which can be used to hedge heating or air conditioning bills, could never survive in an open outcry environment. In an electronic trading environment, weather derivatives can flourish as a group even if no one contract has a high volume.

CURRENT CONDITIONS: CONSOLIDATION AND CRISIS

In the United States and Western Europe, there have been many mergers and substantial consolidation in the futures industry, which is now dominated by five major futures exchanges, each the product of mergers. Three of these firms also include large stock exchanges. In the rest of the world, most of the medium-size and large countries (Japan, India, China, Australia, etc.), continue to have one or more independent futures exchanges or securities exchanges that also offer futures contracts. The five large Western futures exchanges are discussed in the next few paragraphs.

The largest futures exchange in the world today is the CME Group, which is the product of a 2007 merger between the CME and the CBOT and a 2008 merger of the CME Group with NYMEX. The three component exchanges continue to exist with the CME Group serving as parent company. Including NYMEX, the CME Group accounted for 98 percent of the regulated exchange-traded futures volume in the United States in 2007 and 97.2 percent in 2008.[21] Most of the remainder is accounted for by ICE Futures US.[22] In 2008, a group of banks and trading firms created an exchange called ELX [Electronic Liquidity Exchange] Futures. ELX Futures received designation as a contract market (a license to operate a futures exchange in the United States) from the CFTC in May 2009. ELX was expected to list financial futures contracts with low fees to compete with CME Group.

Eurex, as noted, is the product of a merger in 1998 between German DTB and Swiss SOFFEX and trades stocks as well as financial futures. It was the largest futures exchange in the world for several years until the 2007 merger of the CBOT and the CME. Eurex acquired the International Securities Exchange, an electronic equity option market, in 2007. Eurex has not engaged in any additional mergers with other futures exchanges in recent years, but it remains one of the largest futures exchanges. In 2004, Eurex founded a subsidiary, Eurex US that, as noted, attempted to compete with the CBOT Treasury contracts. When those contracts failed, Eurex sold its U.S. subsidiary to the Man Group, which renamed it the U.S. Futures Exchange (USFE). At the end of 2008, USFE closed after a failed attempt to sell it again.

NYSE Euronext is the product of a merger between the New York Stock Exchange and Euronext, which itself is the result of a merger of several Western European exchanges including LIFFE and exchanges in Paris and Amsterdam. At the same time as Eurex was attempting to compete with the CBOT's Treasury note contracts, Euronext listed, unsuccessfully, a Eurodollar futures contract to compete with the CME. NYSE Euronext received CFTC contract market designation for a U.S. futures subsidiary, NYSE LIFFE, in 2008. NYSE LIFFE is expected to list precious metals contracts acquired from the CBOT when the CME Group announced its merger with NYMEX.

ICE, founded as an OTC electronic trading platform in 2000, acquired (as noted) the IPE in 2001 and renamed it ICE Futures Europe. In 2007, ICE attempted to buy the CBOT but ultimately was outbid by the CME. ICE did acquire the New York Board of Trade (NYBOT) in 2007 and renamed it ICE Futures US. NYBOT was itself the product of a merger between the New York Cotton Exchange and the Coffee Sugar and Cocoa Exchange, both with roots dating back to the nineteenth century. ICE also acquired the 121-year-old Winnipeg Commodity Exchange in 2008 and renamed it ICE Futures Canada.

Nasdaq OMX is the product of a merger between Nasdaq Stock Market and OM Group, which operated futures exchanges in the United Kingdom and Sweden. Nasdaq OMX recently acquired the Philadelphia Stock Exchange and its futures subsidiary, the Philadelphia Board of Trade, and had plans to revive PBOT in 2009.

The CBOE remains independent and hoped to demutualize in 2009. Because the CBOE was founded by members of the CBOT, CBOT members had trading rights on the CBOE. The CBOE has been attempting to demutualize for some time but was unable to do so because of a dispute, now resolved, over the CBOT trading rights. The CBOE has also seen some formidable new electronic competitors in recent years: the International Securities Exchange (founded in 1997 and purchased by Eurex in 2007) and the Boston Options Exchange. It is worth noting that it is possible to acquire an option position on one exchange and offset it on another exchange (unlike with most futures contracts).[23]

We conclude with a brief discussion of the possible effects of the events of 2008 on futures and OTC derivatives markets. The year began with an enormous run-up in the prices of most exchange-traded commodities, including grains and energy commodities such as crude oil. Futures trading volume was very high, and high prices were blamed by some on "excessive speculation" in the futures markets and on commodity index traders, which acquire positions designed to replicate commodity price indexes. There were calls for increased regulation and legislation was enacted that subjects the ICE OTC trading platform for energy derivatives to some regulation. Commodity prices mostly declined in the second half of the year, which was characterized by a financial crisis that was blamed in part on OTC derivatives, especially credit derivatives. Credit derivatives, which had grown enormously just in the past few years, were blamed in particular for the near demise of insurance company AIG, one of the largest issuers of credit default swaps. AIG was unable to meet its obligations when a number of credit events occurred and received substantial government assistance. Credit derivatives also played a role in the bankruptcy of Lehman Brothers. Counterparty credit risk again became a concern in the OTC markets. This concern was much more widespread than in 2002, when it was confined to the OTC energy markets. Futures markets performed well during the crisis, and (as had been the case for many decades) there were no defaults or serious problems associated with futures clearinghouses in 2008. There were several attempts under way in early 2009 to set up a clearinghouse for credit default swaps as well as other OTC derivatives such as interest rate swaps. It seems likely that most future credit derivatives transactions and possibly more financial derivatives in general will be cleared in the future. There has also been legislation proposed in the United States to require that all OTC derivatives be traded on regulated futures exchanges, effectively banning the OTC market. Other proposed legislation would require that OTC derivatives be cleared through a regulated clearinghouse without requiring they be traded on an exchange. There has been considerable debate over the extent to which "customized" OTC derivatives should be exempted from any new regulatory requirements and, if so, how "customized derivatives" should be defined. There have been proposals to limit speculative trading in credit derivatives and futures contracts on physical commodities. It seems likely that there will be regulatory changes, but the effect on the markets is still uncertain.

Another possible effect of the crisis is a reduction in futures market volume (or at least volume growth), which as discussed early in this chapter has grown enormously over the last several decades. Industry-wide annual volume increases exceeding 20 percent seemed routine in the United States in the early twenty-first century, but U.S. futures and options volume rose only 4 percent in 2008 over 2007. There were sharp drops in volume in many futures contracts in the latter part of 2008 and in 2009. For example, volume in the CME's interest rate complex declined by more than half in the first four months of 2009 compared to the first four months of 2008. Volume in physical commodities was more or less stable in early 2009 compared to early 2008. Overall futures volume in the United States declined by 28 percent in the first four months of 2009 compared to the first four months of 2008. Equity option volume was down in late 2008 because extremely high volatility made option prices prohibitively expensive for many market participants, but stabilized in early 2009 as volatility declined. A number of major market participants went out of business, merged, or reduced their market exposure during late 2008. These trends appeared to continue in early 2009, but there were signs of improvement in the spring. Because of this, it seems plausible that the rapid growth in futures market volume over the last few decades has, at least temporarily, come to a halt.

ENDNOTES

1. With both futures contracts and forward contracts, the buyer and seller typically enter into a contract for delivery of a commodity at a specified time in the future at a price determined today. Futures contracts have certain elements that distinguish them from forward contracts and make them easy to trade. These elements can include trading on an exchange, standardized contract terms, and the ability to offset a long or short position by entering into an equal and opposite position.
2. The main exceptions were the soybean complex and potatoes.
3. During World War II, the futures industry shrank as many markets shut down temporarily due to price controls.
4. Little is known today about the International Commercial Exchange. The identity of its founders and the exchange's location have been forgotten. Based on FIA's volume reports, the International Commercial Exchange appears to have merged with the New York Produce Exchange before it failed.
5. For example, the CME commissioned Milton Friedman to write a paper entitled "The Need for Futures Markets in Currencies." See Geisst (2002), pp. 182–183.
6. Interest in the futures markets shifted to 5- and 10-year notes because the Treasury stopped issuing 30-year bonds for several years during the first decade of the twenty-first century.

7. Prior the formation of the CBOE, there was some OTC trading in equity options.
8. The specialist system is a type of trading commonly used (especially before the advent of electronic trading) for the exchange trading of securities in which one individual or firm acts as a market maker in a particular security, with the obligation to provide fair and orderly trading in that security by offsetting temporary imbalances in supply and demand by trading for the specialist's own account.
9. The New York Produce Exchange first considered futures on a basket of stocks in the early 1960s. See Falloon (1998).
10. Single stock futures were finally launched in 2002 under joint SEC-CFTC regulation but have suffered from low volumes.
11. CME promotional posters that decorated many Chicago offices in the 1970s and 1980s asked the questions: "How come there's no Peking Duck Exchange?" "How come there's no Moscow Mercantile Exchange?" and "How come there's no Havana Cigar Exchange?" There still is no Peking Duck Exchange, but there are futures exchanges in China, including the Shanghai Futures Exchange and China Financial Futures Exchange. Futures contracts also are traded at the Moscow Interbank Currency Exchange, but there still are no futures exchanges in Cuba.
12. A strip of futures contracts is the purchase or sale of a sequence of contract months in the same futures contract, for example, the March 2010, June 2010, September 2010, and December 2010 Eurodollar futures contracts.
13. Notional principal is a measure of the amount of contracts outstanding (similar to open interest in the futures markets) rather than a measure of trading volume.
14. According to BIS, the outstanding notional principal in exchange-traded equity index derivatives in June 2008 was about $8 trillion. This does not include futures and options on individual equities.
15. The leaded gasoline contract was phased out in favor of an unleaded gasoline contract during the mid-1980s. In the middle of the first decade of the twenty-first century, the unleaded gasoline contract was phased out in favor of a futures contract on reformulated gasoline blendstock for oxygen-blending (RBOB) gasoline. Both of these changes were made to reflect changes in the formulation of gasoline sold in the United States.
16. A crack spread in the futures market is a spread trade between gasoline or heating oil futures and crude oil futures. The process of refining crude oil into gasoline or heating oil is referred to as the crack. The crack spread represents the gross processing margin for refining.
17. EnronOnline was considered profitable. The bankruptcy was caused by losses in other parts of the firm.
18. There had been some instances whereby a larger exchange was able to wrest an existing futures market from a smaller exchange. In each instance, the smaller exchange was unable to create a truly deep and liquid market, but the larger exchange was more successful. Frequently, the larger exchange already listed a related contract, thus facilitating intercommodity spreads (which are easier to execute on a single exchange). During the 1950s, the CBOT, which already had a successful soybean contract, was able to wrest the soybean meal contract from the Memphis Board of Trade and the soybean oil contract from the New York Produce Exchange. Similarly, in 1988, the CBOT wrested the five-year Treasury note futures contract from the Financial Instruments Exchange.
19. Much of the volume in the ICE WTI contract is in spread trades with ICE's Brent crude oil contract. As noted, it is easier to execute intercommodity spreads on a single exchange. The ICE electronic platform has an actively traded natural gas contract (marketed as a swap) that is cash settled based on the NYMEX natural gas futures price.
20. Recall from the beginning of this chapter that U.S. futures volume was fewer than 4 million contracts in 1960.

21. The market share of ICE Futures US increased from 1.5 percent in 2007 to 2.4 percent in 2008.
22. Two tiny futures exchanges that date back to the nineteenth century, the KCBT and the Minneapolis Grain Exchange (which closed its trading floor in December 2008), remain independent. Both exchanges list types of wheat futures that differ from the wheat futures on the CBOT.
23. See Gidel (1999) and the *Futures and Options Factbook* at www.theifm.org. for more information about the various exchanges and their histories. The *Futures and Options Factbook* includes the FIA volume reports which are used extensively in this chapter.

REFERENCES

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, no. 3: 637–654.

Carlton, D. 1984. "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures," *Journal of Futures Markets* 4: 237–271.

Commodity Futures Trading Commission. 1975–2004. *Annual Report.* Washington, DC: Commodity Futures Trading Commission.

Falloon, W. D. 1998. *Market Maker: A Sesquicentennial Look at the Chicago Board of Trade*. Chicago: Board of Trade of the City of Chicago.

Ferguson, N. 2008. *The Ascent of Money: A Financial History of the World*. New York: Penguin.

Ferris, W. G. 1988. *The Grain Traders: The Story of the Chicago Board of Trade.* East Lansing, MI: Michigan State University Press.

Futures and Options Factbook. Washington, DC: Institute For Financial Markets. Online by subscription at www.theifm.org.

Geisst, C. R. 2002. *Wheels of Fortune: The History of Speculation from Scandal to Respectability*. Hoboken, NJ: John Wiley & Sons.

Gidel, S. A. 1999. "100 Years of Futures Trading: From Domestic Agricultural to World Financial," *Futures Industry* 9. Available at www.futureindustry.org/fi-magazine-home.asp?a=607.

Gray, R. W. 1963. "Onions Revisited," *Journal of Farm Economics* 45: 273–276.

Markham, J. 1986. *The History of Commodity Futures Trading and Its Regulation*. New York: Praeger.

Markham, J. 2001. *A Financial History of the United States*. Armonk, NY: M. E. Sharpe.

McLean, B., and P. Eklund. 2003. *The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*. New York: Portfolio Trade.

Melamed, L. 1995. *Leo Melamed on the Markets*. New York: John Wiley & Sons.

Melamed, L., and B. Tamarkin. 1993. *Escape to the Futures*. New York: John Wiley & Sons.

Steinharr, A. 1998. *Derivatives, the Wild Beast of Finance*. New York: John Wiley & Sons.

Tamarkin, B. 1993. *The MERC: The Emergence of a Global Financial Powerhouse*. New York: HarperBusiness.

ABOUT THE AUTHOR

Michael A. Penick has worked for the CFTC since 1991 and for its Office of the Chief Economist since 2002. He has worked with the Division of Enforcement on numerous market manipulation investigations, including Amaranth and Enron. Research interests include the economic history of futures trading and regulation. He created and maintains the CFTC History pages on the CFTC Web site and maintains the CFTC glossary. Prior to 2002, he reviewed new futures contracts and rule amendments for the CFTC's Product Review Group. Previously, he was a member of the Chicago Board of Trade, where he traded options on Treasury bond futures and stock index futures and a stock market analyst for a market maker firm on the Chicago Board Options Exchange. He received an MA in economics from Stanford University in 1984 and a BA in mathematical methods in the social sciences and economics from Northwestern University in 1982.

CHAPTER 18

Derivatives Markets Intermediaries

Brokers, Dealers, Pools, and Funds

JAMES L. CARLEY
Associate Regional Director, Securities and Exchange Commission

In derivatives markets, as in securities markets, there are different types of market intermediaries. These fall into two broad categories: those who provide transaction execution services by accepting market orders from other persons and/or by executing market transactions on their behalf and those who perform a money management role by providing advice on derivatives transactions to other persons or by directly managing asset portfolios for such other persons. Neither by economic function nor by government decree are those categories or subcategories mutually exclusive. A number of firms perform multiple intermediary functions both within and across market sectors. These intermediaries may operate in public markets, such as through the organized exchanges across the globe on which many standardized futures and option contracts are traded, and/or in the global markets for privately negotiated (also called over-the-counter or OTC) derivatives transactions, such as that for the bilateral agreements commonly referred to as swaps.

Many of these firms are subject, with respect to at least some aspects of their operations, to regulation and supervision by one or more instrumentalities of the federal government and/or industry self-regulatory organizations (SROs). Other firms are not currently subject to direct regulation or supervision, particularly with respect to their OTC activities, but recent market events and economic conditions have triggered increased discussion of the potential for change in this regard.

The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The views expressed herein are those of the author and do not necessarily reflect the views of the Commission or of the author's colleagues upon the staff of the Commission.

INTERMEDIARIES FOR EXCHANGE-TRADED DERIVATIVES

In the United States, the Commodity Futures Trading Commission (CFTC) is the primary functional regulator of entities that perform, with respect to many but not all types of derivatives, one or more of the functions just described.[1] While the CFTC generally has exclusive jurisdiction over futures market activities, it does not have exclusive jurisdiction over the entities that participate therein. Many firms supervised by the CFTC with respect to their futures market activities are also regulated by one or more other functional regulators on the basis of their activities in other market sectors.

For example, a number of firms provide brokerage services for both derivatives and securities transactions; the latter subjects them to the jurisdiction of the U.S. Securities and Exchange Commission (SEC) as well as that of the CFTC. Moreover, the ownership or control of some firms that are active in the futures markets—for example, as part of a bank holding company—may subject them to the jurisdiction of a bank supervisor such as the Board of Governors of the Federal Reserve System. However, in the past that authority has not been exercised directly very often, with greater deference to the respective functional regulator being the norm. In some other nations, entities that are involved in one or more aspects of the financial markets may fall under the actively exercised jurisdiction of a consolidated financial supervisor, such as the United Kingdom's Financial Services Authority (FSA), which oversees the entire financial services sector, from banking to insurance to trading in equities and derivatives.[2]

In addition to government oversight, market intermediaries also also be subject to direct examination of their financial condition, surveillance of their market activities, and member supervision by one or more nongovernmental SROs. There is currently one SRO, the National Futures Association, Inc. (NFA), in which every such intermediary operating in the United States must maintain membership.[3] However, as will be discussed, the NFA is not the only entity in the United States carrying SRO responsibilities with respect to futures intermediaries; the futures exchanges also have obligations to perform certain SRO functions with respect to firms that are trading and clearing members thereof.

The legislation controlling futures markets in the United States is the Commodity Exchange Act (the act or CEA), passed in 1936.[4] Under the CEA, the regulatory framework of the CFTC, and the rules of the NFA and other SROs, the four functional types of intermediaries discussed at the outset are referred to, respectively, as introducing brokers (IBs), futures commission merchants (FCMs), commodity pool operators (CPOs), and commodity trading advisors (CTAs).

Generally, it is the performance of the respective functions of an IB, FCM, CPO, or CTA for or on behalf of U.S. customers (i.e., persons located within the United States) that is determinative of a particular firm's registration requirements with the CFTC and not the domicile of that firm nor the location of the futures exchange on which transactions may be executed. Thus, a firm located in London that executes transactions on a German futures exchange for U.S.-domiciled persons generally would be deemed to be an FCM and subject to the CFTC's jurisdiction.

As noted, a particular intermediary may perform more than one of these four basic types of functions and thus be required to register with the CFTC in multiple capacities. For example, a number of firms manage pooled investment vehicles that enter into futures contracts (commodity pools) in a manner that constitutes both the operation of, and the provision of trading advice to, such pools and are therefore registered with the CFTC as both CPOs and CTAs. In addition, a particular firm may engage in activities outside the futures markets that subject it to the jurisdiction of another federal financial regulator. As noted earlier, a number of financial firms engage in brokerage activities with respect to securities and futures and are therefore registered both as broker-dealers (BDs) with the SEC and as FCMs with the CFTC.

Providers of Trade Execution Services

Firms that provide trade execution services to futures customers fall into several broad categories, divided primarily on whether they handle cash and other customer assets or merely process orders and whether they execute customer transactions directly on a futures exchange or do so through another firm.

Futures Commission Merchants

From a regulatory perspective, an FCM is defined as any individual or organization that both: (i) solicits or accepts orders to buy or sell futures contracts or commodity options (or, options based on futures); and also (ii) accepts from customers cash, securities, or other property to support such orders (i.e., to serve as margin or performance bond ensuring the fulfillment of contractual obligations). Any person that engages in both of these activities must register with the CFTC, and there are no exemptions to that requirement. The regulatory regime for FCMs (enforced through audits and examinations by the CFTC, NFA, and the staffs of the major exchanges and clearinghouses) focuses on record keeping, customer disclosures, and financial and risk reporting, in addition to minimum capital requirements.

From a clearing system perspective, there is a major distinction between clearing FCMs, which are members of one or more futures exchanges and the affiliated clearing organization(s), and nonclearing FCMs. A nonclearing FCM solicits and accepts customer orders and the customer funds to margin the resulting futures positions. However, because it does not have the ability to execute transactions directly, it must establish an omnibus account (or a fully disclosed arrangement) with a clearing FCM to effect those transactions. Clearing FCMs tend to be larger, corporate entities that execute futures transactions for customers through floor brokers that may be employed as staff of the FCM and/or through independent floor brokers. (Those persons who execute futures transactions only for themselves are known as floor traders.) Nonclearing FCMs are often smaller, and sometimes family-run, businesses.

Due to competition and consolidation, the number of FCMs continues to decline steadily. In contrast to the roughly 400 FCMs registered in the mid-1980s, there were only 149 FCMs registered with the CFTC by 2008. Approximately 40 percent of these FCMs are also registered as BDs, and these dually registered BD/FCMs are responsible for more than 80 percent of futures customer business, as measured by the customer funds on deposit to margin those transactions. The balance sheets of these 60 dually registered firms represents 96 percent of the more than $100 billion in capital of all FCM firms.[5] The largest of these brokers of both securities and futures include some of the most widely recognized firms on Wall Street.

There are, however, important distinctions between futures brokerage and securities brokerage activities, one of which is the much greater extent to which transactions in futures, as derivatives, are leveraged relative to securities. Whereas the purchaser of a security generally is prohibited from posting less than half the value of such security, futures contracts permit both physical commodities and financial positions to be controlled with margins of less than 10 percent. This means that the parties that intermediate these futures transactions—that is, as between those persons taking long or short futures contract positions on one side and the exchange and its affiliated clearinghouse on the other—provide more than mere transaction execution services; these firms play key credit support and risk management roles in the futures markets.

Although few in number compared to the thousands of registered BDs, the roughly 150 FCMs (including BD/FCMs) hold more than $200 billion in customer funds to margin futures positions on U.S. and foreign boards of trade. Because the risk margin (or performance bond) requirement for a futures contract is typically in the range of 3 to 7 percent of the contract's notional value, this level of customer funds can support open contracts with an aggregate notional value in the trillions. And, while the number of FCMs may be declining, the volume of futures transactions that they execute continues to grow at a double-digit rate each year.[6] For example, global trading volumes in 2007 grew overall by more than 28 percent versus the prior year. For North American exchanges, the growth was almost 33 percent. Growth was consistently high across interest rate products, equity index contracts, and the traditional physical derivatives, such as energy and agricultural commodities.

Derivatives contracts are an economically efficient tool because they permit market participants to manage the full risk of an exposure without having to commit capital in the entire amount of the notional (or face) value of that exposure. If that were not the case, and if each party to the contract were required to post collateral equal to the notional value of a derivative contract, entering into such contracts would have no advantage over taking a direct position in the spot market (which, of course, is not even possible with all of the things that may underlie the expanding universe of derivatives). One way to achieve capital efficiency while also ensuring the reliability of the inherent promises of timely payment that make these risk management tools effective is to require the initial posting of performance bond deposits and to enforce frequent mark-to-market valuation of positions so that the gain or loss on a particular position is recognized and settled rather than allowed to accumulate until expiration of the contract. For example, U.S. FCMs hold more than $200 billion to margin futures positions on U.S. and overseas exchanges whose aggregate notional value is in the trillions. This extensive level of risk management activity is possible because daily (and sometimes intraday) mark-to-market margining allows additional performance bond deposits to be called forth quickly from futures brokers in response to changes in the market.

In other words, because additional deposits can be called forth from a customer (or from such customer's broker, should the customer default) and because a customer's failure to promptly post additional deposits with the FCM in response to such a call can result in a forfeiture of that customer's positions, risk margin requirements can be based on the price volatility expected to occur over a very short time horizon (typically one day) with a very high confidence level (typically 99 percent) and still be both prudent and efficient from a capital commitment perspective. (The amount of risk margin that would be required to support positions with less frequent mark-to-market margining would make futures uneconomic as risk management tools.)

Thus, the customer monies deposited with an FCM represent primarily the margin required to support the inherent promises of those customers to make timely payments under derivatives contracts (plus, e.g., realized but undistributed customer profits). There is, accordingly, a strong interest on the part of both the clearing system and the federal financial regulators who supervise it in seeing that controls are designed and maintained to help ensure that such monies are held in a manner that provides immediate accessibility to the clearinghouse in the event they are needed to cure a default and avoid any contagion effect on other market participants.

Thus, for example, key provisions of the bankruptcy code and the CFTC regulations seek to provide for unambiguous application of customer funds in the event of an FCM's insolvency. In fact, the Commodity Exchange Act initially permitted customer funds to be invested only in the most traditionally stable value government securities. However, CFTC regulations now permit a somewhat broader range of investments, including such vehicles as money market mutual funds, commercial paper, high-rated corporate debt, and even certain repurchase agreements and reverse repurchase agreements. (It should be noted that neither the law nor regulations require that earnings on such investments of customer funds be returned by the FCM to customers; however, these investment earnings may provide an indirect benefit to customers in strengthening the capital position of the FCM, particularly in an era when competition among firms has reduced transaction-based revenues for brokerage firms.)

Liquid, stable-value investment vehicles for customer margin deposits are an important systemic protection because clearinghouse rules consistently permit the application by the clearinghouse of the margin deposits of nondefaulting customers of an FCM to cover the obligation of a defaulting customer of that same FCM. In other words, futures customers at the same FCM, notwithstanding legal and regulatory requirements that an FCM hold separate and apart from its own funds (or segregate) the assets of its customers as a group, effectively mutualize the default risks of their fellow customers within that group.

In addition, even at a 99 percent confidence level, daily price movements on a futures contract can be expected to exceed the risk margin level several times during a year of roughly 250 trading days, so there is a need to backstop the margin system with an appropriate level of FCM capital. In recent years, the regulatory requirements for minimum capital have been revised to now be based primarily on the risk margin required by futures exchanges with respect to the positions carried by an FCM for its customers. This modern risk-based approach to capital requirements replaced a prior focus on the amount of customer funds held by the FCM and represented an important advance in the regulatory structure.

Among other things, the prior framework created a disincentive for an FCM to require additional deposits from a customer in whose creditworthiness the FCM might have less confidence and from whom the FCM might therefore choose to require additional deposits. Today, an FCM's minimum capital requirement generally is based on the sum of 8 percent of the margin required by the net customer positions it carries and 4 percent of its proprietary positions. Thus, a firm that carries a relatively balanced book of customer business will face lower regulatory capital requirements than one that has a significant net exposure, long or short, arising from its own or its customers' positions. Capital requirements now reflect incentives consistent with more responsiveness to changes in market volatility.

Just as aggregate customer margin deposits may appear small relative to the notional values of the futures contracts they support, so might regulatorily mandated FCM capital levels appear small relative to the margin deposits for which those FCMs are responsible to customers. Only about half of all FCMs, fewer than 80 firms, actually hold margin deposits and other customer assets, in an aggregate amount of more than $200 billion. While they together maintained actual capital of almost $90 billion in 2008, these firms were in fact only required by regulation to maintain less than $20 billion, or approximately 8 percent of their combined liability to futures customers.

It should be noted, however, that the CFTC does monitor material changes in capital levels and events such as significant withdrawals or other changes do draw supervisory attention. Indeed, the CFTC has been innovative in utilizing existing information to enhance its financial and risk surveillance of firms, even in advance of capital deterioration. The agency's systems combine (i) trading activity information that is required to be submitted by futures traders and their brokers to the CFTC to assist the agency in its efforts to detect and deter market manipulation and that allows its economists to assess the size and extent of a particular person's market exposure even across different accounts and different brokers with (ii) financial condition information about the brokers from the Form 1—FR and FOCUS reports that they must file. This provides the agency with a unique view of market and counterparty risks and the ability to assess the capacity of those brokers to handle such risks.[7]

The capital of an FCM is an important element of protection for both its own futures customers and for the integrity of the clearing system as a whole. The CFTC has clearly articulated its position that an FCM, in a circumstance where a customer default causes such a forfeiture of assets equitably owed to nondefaulting customers, has an affirmative obligation to top up the segregated funds of nondefaulting customers.[8] Its primary means of doing so, of course, will be through its own capital. Their commitment of proprietary capital and the risk of loss incumbent therewith give FCMs powerful incentives to perform effective risk management with respect to the derivatives positions they carry for their customers. An FCM that does not effectively manage such risks may find itself out of business very quickly.

Unfortunately, the consequences of an FCM's failure to manage its risk exposures properly might not be limited to the FCM itself or even to its customers. As noted, nondefaulting customers of an FCM that fails to prevent a default by one customer can be affected immediately should the customer's segregated funds account at the clearinghouse be forfeited in the event the FCM's capital is inadequate to cover the default. However, more severe losses can cascade systemically if guaranty funds deposited by other clearing members must be called pursuant to clearinghouse rules and if, as also provided by clearinghouse rules, other firms are required to meet a special call for additional contributions. And, of course, in any such situation, the reputational capital of the exchanges, clearinghouses, and member firms can be jeopardized.

Accordingly, there are strong economic incentives for exchanges and clearinghouses to perform effective self-supervision and for the FCMs to support and cooperate therewith. In addition to the audit staffs maintained by the CFTC and the NFA in various cities, the exchanges themselves have maintained audit and financial surveillance functions. To avoid duplication and redundancy in how the financial condition of member firms is conducted, the futures SROs (including the NFA and both those exchanges that maintain audit staffs and those that do not) have for several decades coordinated their efforts through a Joint Audit Committee (JAC), pursuant to CFTC rules permitting such cooperation.

Traditionally, the examination of FCMs that did not maintain membership at any exchange (i.e., nonclearing firms) would be performed by the NFA. Examination of an FCM that was a member of only one exchange usually would be performed by the staff of that exchange. Examination of each of the largest FCMs, generally including all or most of the BD/FCMs, which typically would be members of multiple exchanges, was assigned to one the major exchanges, and the JAC agreement provided for information sharing among the other SROs so that each exchange at which such firm was a member would have access thereto without having to perform duplicative examination work. (The examinations of FCMs, whether performed by the NFA or one of the exchange SROs, do cover sales practices and other customer protection and regulatory compliance issues as well as an assessment of financial condition and capital compliance.) The JAC agreement has not been amended since 1984. With significant consolidation having occurred among exchanges, the CFTC has recently sought public comment on a major proposed amendment thereto.[9] The CFTC, in an oversight regulator role, monitors the functioning of the JAC process and the individual performance by each SRO of its examination and other responsibilities and also retains the authority to examine any FCM or other CFTC-registered entity at any time.

Introducing Brokers

Introducing brokers resemble FCMs in their solicitation and acceptance of customer orders for futures contracts to be entered into on a futures exchange but, unlike an FCM, an IB does not accept customer assets (neither cash nor securities) as performance bond against such contracts nor does it extend credit in lieu of such deposits. Any person or organization that engages in such solicitation and acceptance of customer orders must register with the CFTC; however, regulatory exemptions exist for persons separately registered as CPOs or CTAs (to be discussed). Because they generally do not hold customer assets, the regulatory capital requirements imposed on IBs are much lower than those for FCMs, and some IBs can satisfy these requirements by being guaranteed by an FCMs (called GIBs), a status that does not imply significant regulatory distinctions for the GIBs themselves but may have certain implications for those guarantor FCMs, including higher minimum capital requirements of those FCMs and potential joint and several liability to customers with respect to sales practices and other activities. IBs tend to have a very small scale of operations.

Providers of Money Management Services

Whereas the number and types of firms that provide trade execution services to futures customers are relatively finite, the number and variety of persons and firms that provide portfolio management and/or trading advice (and, thus, may fit the statutory definitions of CPO and/or CTA) are immense. These may include heavily regulated entities such as mutual funds, small personal advisors, providers of certain online services, and, perhaps most notably, a great many entities among the large and diverse population of pooled investment vehicles that are exempt from SEC registration and often known as hedge funds (although such term is nowhere defined in the act or in CFTC rules).

Commodity Pool Operators

Under the CFTC rules, a commodity pool has been defined broadly as "an investment trust, syndicate, or similar form of enterprise operated *for the purpose of* trading in commodity interests." This could cover a wide variety of pooled investment vehicles, some of them with portfolios focused heavily or exclusively on derivatives but others with few futures positions, such as a stock mutual fund that may use stock index futures to hedge its equity holdings. A person operating a commodity pool normally must be registered with the CFTC and supervised by the NFA. (Unlike mutual funds that must be registered as investment companies with the SEC, commodity pools themselves are not required to be registered with the CFTC.)

The regulatory regime with respect to CPOs has been described as the "four Rs," representing: registration, record keeping, (risk) disclosure, and reporting. Unless it qualifies for one of the various exclusions or exemptions provided by the CFTC rules, a CPO must register with the CFTC (and become a member of the NFA), adhere to certain bookkeeping and record retention requirements, provide prescribed disclosures to prospective investors, and make certain periodic reports to investors (and the NFA). Unlike the regulatory regime with respect to FCMs, there is no general minimum capital or other financial integrity requirement for CPOs.

The CFTC has provided certain exclusions and exemptions to the registration and other regulatory requirements for certain categories of persons notwithstanding their operation of pooled investment vehicles that would be deemed to be commodity pools, primarily those that are otherwise regulated (such as mutual funds, insurance companies, banks, and pension plans operators) and those that operate pools in which futures trading represents less than a *de minimis* level of portfolio activity and/or in which participation is restricted to sophisticated investors.

For example, one exclusion formerly required that a pool restrict futures trading to bona fide hedging activity and to refrain from committing more than 5 percent of portfolio assets to the trading of commodity interests. However, the CFTC has since relaxed these restrictions on the purpose and scope of futures trading by qualifying otherwise-regulated entities (such as mutual funds and insurance companies), which may now engage more extensively in futures trading with no ongoing regulatory interaction with the CFTC or NFA. (However, even these excluded entities remain subject to special calls for information from the CFTC and to the antifraud and large-trader reporting requirements of the Commodity Exchange Act.)

The rules with respect to pools in which participation is restricted to sophisticated investors and/or which engage in limited futures trading activity do not afford the complete regulatory exclusion granted to otherwise-regulated entities, but they do provide key forms of exemptive relief. Thus, for example, one CFTC rule has long provided a "light" version of regulation by permitting the operators of pools in which participation is restricted to persons that meet certain regulatorily defined thresholds of being highly qualified (which consider both income and wealth and which differ for natural persons and legal entities) to make less detailed (but nonetheless not misleading) customer disclosures and to make and file less frequent reports. Even though regulated more lightly, such CPOs must nonetheless register with the CFTC, while excluded (otherwise regulated) entities do not have to register.

However, more recently adopted rules also provide exemptions even from the registration requirement for those CPOs operating commodity pools that either: (i) restrict participation to highly qualified investors and limit futures trading to a *de minimis* level (less than 5 percent of assets committed thereto); or (ii) restrict participation to much more highly qualified investors (but with no limitation on the extent of futures trading). Although exempt from registration with the CFTC, such CPOs do remain subject to special calls for information from the CFTC, to the antifraud and large-trader reporting requirements of the Commodity Exchange Act, and also to certain regulatory requirements such as those with respect to the accurate advertising of past performance.

Commodity Trading Advisors

CTAs provide advice about futures transactions and markets. Some may exercise, through powers of attorney, actual discretion over individual client investments, whether directly held futures positions or interests in commodity pools. Other CTAs may provide advice and trading direction to one or more commodity pools but not to individual natural person clients (and a single commodity pool may be advised by multiple CTAs).

Although many persons and firms are dually registered as both CPOs and CTAs, and thus subject to the full set of the four Rs described earlier, many others remain strictly in an advisory role. These CTAs are subject to registration, recordkeeping, and disclosure requirements but not to periodic reporting requirements. (Note, however, that clients receive trade confirmations and account statements from the FCM that carries their futures positions.)

Supervision of CPOs and CTAs

Unlike FCMs, the thousands of CPOs and CTAs have not traditionally been examined by any of the exchange SROs but, rather, by the NFA, with oversight of those efforts by the CFTC. The CFTC has delegated to the NFA the function of receiving the reports and customer disclosures required to be filed by CPOs and CTAs but, as with FCMs, retains the authority to access these records and/or to examine the firms directly.

INTERMEDIARIES FOR OTC DERIVATIVES

While exchange-traded derivatives provide a number of important benefits for those seeking effective risk management tools, such as providing greater liquidity and reduced counterparty exposure, the need for contract fungibility in exchange-traded and centrally cleared environments may make this type of derivative a suboptimal tool for certain users or particular circumstances. In contrast, privately negotiated derivatives, or swaps, provide the opportunity for counterparties to tailor the terms of a derivative contract to their particular needs. Thus, in addition to those activities of derivatives intermediaries that involve exchange-traded futures and options and are supervised directly by government regulators and SROs, a number of these intermediaries or their affiliates participate in the OTC derivatives markets. These activities traditionally have not been subject to much, if any, regulatory scrutiny, and currently there is no formal SRO in this sector.[10]

Swap Brokers

While the bilateral nature of swaps makes them advantageous to those seeking a more individually tailored risk management tool, it also presents a major challenge to such derivatives users: Finding an available counterparty with appropriately converse risk management goals and to whose credit risk one is willing to expose oneself is difficult and potentially very costly. A swap broker is an intermediary that maintains a client base of firms to which it provides matchmaking assistance. Once a match is made and the prospective counterparties are properly introduced, the swap broker also may provide assistance in negotiating and documenting the transaction. These facilitation services typically are compensated through fees charged to one or both counterparties.

Thus, in contrast to a futures commission merchant that, through its implicit guaranty of its customer's transaction, acts in a centrally cleared market as a risk intermediary, a swap broker is not a party to the transactions it facilitates and acts strictly as an information intermediary. Ostensibly, the overtures made by a broker on behalf of its client to potential counterparties, or to another broker in search of such potential counterparties, are made anonymously with respect to the client's identity. In reality, however, and particularly in market environments characterized by heightened concern over counterparty credit risk, a broker may well demand greater information about a potential counterparty before moving ahead with an introduction.

Swap Dealers

Even with the assistance of a matchmaking swap broker, it remains uncertain whether a willing and qualified counterparty can be found when a client seeks to enter into a swap to accomplish its risk management goals. Some financial firms, therefore, are willing to stand ready to interpose themselves as counterparties opposite clients. In doing so, they become risk intermediaries but then may seek to reduce some or all of their resulting risk exposures through offsetting swap transactions with one or more third parties and/or by taking risk mitigating positions in exchange-traded futures. Swap dealers can derive significant profit from the spreads in such transactions.

Interdealer Brokers

Nonetheless, a swap dealer itself, even with its superior knowledge of the markets and participants, may face significant difficulty and expense in locating appropriate and available counterparties to offset risks that result from it having stepped forward as a swap counterparty to its client, risks that may not always be manageable through exchanged-traded futures. A natural next step has thus become the emergence of the interdealer broker (IDB). An IDB is a specialist intermediary among swap dealers but is not a centralized counterparty; that is, it fills a role in some ways more like that of a trading facility than of a clearinghouse. An IDB performs a transaction execution role and also may provide price discovery, at least for the swap dealers with whom it does business, if not to the broader marketplace. Analogous to the role of a swap broker between two end user counterparties, the IDB is an informational intermediary, facilitating the transaction between two swap dealers but not becoming a party thereto or exposing itself to the creditworthiness of either dealer.

Next Step: Clearinghouse(s) for Swaps

With recent market events and economic conditions, there has arisen enormous interest among both market participants and financial regulators in expanding the availability of centralized clearing for swaps transactions.[11] Indeed, the President's Working Group on Financial Markets (PWG), working with the Comptroller of the Currency and the Federal Reserve Bank of New York, has actively encouraged the exploration of ways to bring enhanced transparency, integrity, risk management, and regulatory cooperation to the OTC derivatives markets.[12]

In late 2008, the Federal Reserve Board of Governors, the CFTC, and the SEC entered into a memorandum of understanding (MOU) memorializing their intent to "cooperate, coordinate and share information . . . in carrying out their respective responsibilities and exercising their respective authorities with regard to Central Counterparties for credit default swaps" (or CDS), which are defined therein as:

> *any entity that is engaged in the business of interposing itself, either* through a guarantee or as a principal*, between counterparties in providing clearing and settlement facilities for post-trade processing of OTC credit default swap transactions or both OTC and exchange-traded credit default swap transactions [and] does* not *include inter-dealer brokers. [Emphasis added]*[13]

While a number of efforts in this direction had already been under way, recent market events have provided an incentive, both for the industry and for the regulators, to accelerate development and approval thereof. According to press accounts, at least three major exchanges in the United States and one in Europe were working actively in late 2008 to initiate clearing operations for credit default swaps by the end of the year.[14] The PWG's own announcement of the MOU stated:

> *[CCPs can] reduce the systemic risk associated with counterparty credit exposures [as well as] help facilitate greater market transparency and be a catalyst for a more competitive trading environment that includes exchange trading of CDS. . . . The relevant regulatory authorities are assessing these [CCP] proposals by conducting on-site reviews [and] expect to proceed toward regulatory approvals and/or exemptions expeditiously [anticipating] that one or more [CCPs] will commence operations before the end of 2008.*

And, indeed, in December 2008, the SEC provided exemptions allowing such an arrangement to operate as a central counterparty for credit default swaps, stating that the action—which was taken in consultation with the Federal Reserve Board of Governors, the New York Fed, the CFTC, and the U.K.'s FSA—was "an important step in stabilizing financial markets by reducing counterparty risk and helping to promote efficiency in the credit default swap market."[15]

ENDNOTES

1. See www.cftc.gov for direct access to CFTC rules and interpretive guidance, a glossary and primer, market and trading data, and other information. A section entitled "Industry Oversight" has links to more specific web pages discussing the different types of derivatives market intermediaries.
2. See www.fsa.gov.uk. Discussion of such consolidation among the various U.S. functional regulators has increased in the wake of recent market events and economic conditions.
3. See www.nfa.futures.org. The NFA web site provides various publications on futures and options trading as well as investor protection guidance. It also provides access to the Background Affiliation Status Information Center (BASIC) database, which allows one to view registration and membership information about firms and other persons in the futures industry as well as futures-related regulatory and nonregulatory actions taken by the CFTC, NFA, or another SRO against such persons.
4. Codified in the United States Code at 7 U.S.C. Section 1, et seq. It was enacted on June 15, 1936, to replace the Grain Futures Act of 1922.
5. The CFTC provides through its web site, on a monthly basis, financial information submitted by FCMs. The CFTC makes this information publicly available in a spreadsheet format showing for each FCM: its designated self-regulatory organization, its minimum required and actual capital levels, and the amount of funds it holds for customers with respect to positions on domestic and overseas exchanges.
6. The Futures Industry Association publishes on its web site at www.futuresindustry.org a variety of volume statistics about domestic and global trading volumes in commodity futures and options. Articles on both market and regulatory issues can be found at www.futuresindustry.org/fi-magazine-home.asp.
7. The CFTC's FY2008 President's Budget and Performance Plan submission, available at www.cftc.gov/aboutthecftc/2008budgetperf-txt.html, describes its systems: "With the establishment of a financial surveillance unit, the [CFTC] has an enhanced capability to monitor market information, evaluate the impact of market moves on the financial integrity of market participants, and anticipate and act upon indications of financial difficulty. This capability is built upon the implementation and use of the new FSIS component systems, including the RSR Express system that compiles FCM financial statements, the SPARK system that utilizes large trader information to allow the tracking of risk at market, firm, and account levels, and the SPAN Risk Manager system that will permit 'what if' analyses."
8. See CFTC Letter No. 00-106, dated November 22, 2000, promulgated by the Division of Trading & Markets, which can be found at www.cftc.gov/tm/letters/00letters/tm00-106.htm.
9. See 73 FR 52832 (September 11, 2008).
10. It should be noted, however, that the International Swaps and Derivatives Association, Inc. (ISDA) has developed and promulgated a widely used Master Agreement that provides standardization for many of the ancillary terms and conditions of a swap agreement while leaving the individual counterparties with flexibility to negotiate the core financial and risk provisions that best fit their particular needs. ISDA has also been

involved with work on netting and collateral agreements. See www.isda.org. As an indication of how widely used the ISDA master agreement is, recruiting firms in New York and London actively seek candidates for the position of ISDA Master Negotiator.

11. There have been some targeted efforts in providing clearing services to particular segments of the OTC derivatives markets, such as the separate efforts by the New York Mercantile Exchange and the InterContinental Exchange in Atlanta to provide clearing for OTC energy contracts.
12. The PWG was formed in response to market events of October 1987 and consists of the secretary of the Treasury and the chairmen of the CFTC, the Federal Reserve Board of Governors, and the SEC. The PWG does not have its own web site, but information about its activities and reports can be found through the Treasury web site at www.treas.gov. For example, the PWG's report on Long Term Capital Management can be found at www.treas.gov/press/releases/reports/hedgfund.pdf.
13. See www.ustreas.gov/press/releases/hp1272.htm.
14. See www.bloomberg.com/apps/news?pid=20601103&sid=a8ZteOko.5KI&refer=news.
15. See www.sec.gov/news/press/2008/2008-303.htm.

REFERENCES

In addition to the web sites noted, each of which can provide to students of the derivatives markets a wealth of information on the history and current state of the markets, these texts provide both fundamentals and more advanced discussions of the topics introduced herein:

Kolb, R. W., and J. A. Overdahl. 2003. *Financial Derivatives*, 3rd ed. Hoboken, NJ: John Wiley and Sons, Inc.

Kolb, R. W., and J. A. Overdahl. 2006. *Understanding Futures Markets*, 6th ed. Malden, MA: Blackwell Publishing.

Kolb, R. W., and J. A. Overdahl. 2007. *Futures, Options and Swaps*, 5th ed. Malden, MA: Blackwell Publishing.

ABOUT THE AUTHOR

James L. Carley is the associate regional director for examinations in the Atlanta Regional Office of the U.S. Securities and Exchange Commission, where he leads a team of accounting, finance, and legal professionals in the examination of broker-dealers, investment companies, investment advisors, and transfer agents located across the Southeast. He joined the SEC in 2006 after seven years in various roles with the U.S. Commodity Futures Trading Commission, including leading the supervision of derivatives clearinghouses, futures brokers, and commodity pool operators and trading advisors. Before entering public service, Mr. Carley served in financial management roles with the GTE Corporation and USAirways and practiced law in the area of aviation finance with the Washington, D.C., firm of Zuckert, Scoutt & Rasenberger. Mr. Carley holds a BS in economics, a JD from George Mason University, and an MBA from the University of Texas at Austin.

CHAPTER 19

Clearing and Settlement

JAMES T. MOSER
Deputy Chief Economist, U. S. Commodity Futures Trading Commission

DAVID REIFFEN
Senior Economist, U. S. Commodity Futures Trading Commission

INTRODUCTION

Two important differences between futures contracts traded on organized exchanges and over-the-counter (OTC) forward contracts relate to the likelihood of counterparty default and the ability of traders to unwind their positions prior to expiry (i.e., return to a zero net position from an initial long or short position). The ease of unwinding and the lower likelihood of default associated with exchange-traded contracts, in turn, relate to the fact that exchange-traded contracts are cleared and settled by a clearinghouse.[1] This chapter explains how clearinghouse rules enable exchanges to provide contracts containing these features. We also discuss differences in the rules established by different clearinghouses and relate these differences to the characteristics of the underlying products being traded. We then discuss some recent proposals to change the structure of clearing for futures and options on futures.

FUNCTIONS OF CLEARINGHOUSES

Counterparty default is an important concern in any trading environment. In general, this is because defaults result in deadweight losses in the form of costs associated with collecting against defaulting parties as well as costs incurred in restoring positions. Addressing these problems is the province of the clearinghouses that make the arrangements needed to clear and settle transactions. This section broadly describes these arrangements for two transaction categories: those requiring immediate performance[2] and those specifying deferred performance.

We have benefited from the comments of Pat Fishe, Mark Manning, John McPartland, Jim Overdahl, Daniela Russo, and Kirsten Soneson. The views expressed in this paper are those of the authors and should not be viewed as those of the Commodity Futures Trading Commission or any of the commissioners.

Contracts for Immediate Performance

Normally construed, transactions such as security purchases consist of a party agreeing to pay money in exchange for another party agreeing to deliver an asset. Conceivably, transactions so described could be completed on the spot with an exchange of cash for securities; hence, these markets are referred to as "spot" or "cash" markets. Given a certain degree of heterogeneity in the cash and securities handling practices of market participants, however, it is typically more practical for the transacting parties to agree to allow specialist firms to handle the actual transfer details.

The specialist firms that operate the systems that settle transactions do so by completing the details necessary to effect trades. Hence, clearing entails identifying and reconciling the obligations created by the trades of two counterparties: What funds are being transferred, from and to which accounts, which securities are being delivered, as well as where from and where to? Construction of a computer program provides a useful analogy. As the series of steps needed to effect the trade are identified, they are compiled as processing instructions, and those instructions are then executed.[3] When the processing steps are acknowledged as having been complete, the trade is said to have been settled.

Since the steps involved in processing transactions tend to be characterized by scale economies, it is not surprising that the processing of transactions in every security on an exchange typically is performed by a single specialized firm called a clearinghouse. Some more specialized clearinghouses go beyond this processing function and provide guarantees that agreements made between transactors will be completed. These guarantees are accomplished by the clearinghouse accepting the liabilities of the contracts between clearinghouse members that it clears. Hence, should any clearing member default on its obligations to any other clearing member(s), the clearinghouse stands ready to fulfill the obligations of the defaulting member. Assumption of obligations entails use of financial resources obtained from nondefaulting (or surviving) clearing members. This use of a shared pool of resources accomplishes loss mutualization. Many clearinghouse features are direct consequences of the presence of mutualized loss. For our purposes, two features are particularly relevant:

1. Because each clearing member's loss exposure depends on the financial status of other clearing members, clearinghouses establish financial standards for their members.
2. Mutualization covers only default by one clearing member to another, as will be detailed; the obligations of a defaulting member to those outside the clearinghouse are not covered.

While clearinghouses guarantee payments between clearing members, clearing members serve a similar role in guaranteeing promises made to other parties, such as exchange members who are not clearing members and nonmember traders. The net effect of this contractual arrangement is that the promises made by each trade counterparty (e.g., that she will deliver the shares, that he will deliver the cash) is backed up by the clearing system, so that neither party to a transaction need know anything about the financial condition of his or her counterparty. As such, trading decisions can be based solely on prices bid and offered. Thus, aside from the economies obtained by having specialists handling the details of payment and delivery, an advantage of this system is that it enables anonymous trading. Such anonymity can be valuable to buyers and sellers. For example, for transactions affecting corporate control, traders may prefer not to be identified. Alternately, were payments and deliveries made directly between buyers and sellers, their identities would be known to outsiders. Routing of payments and deliveries through a clearinghouse amounts to redacting the buy-side and sell-side identities from every transaction. Thus, for example, a large fund needing to improve its diversification can sell off securities without fearing the adverse impact of the trade becoming viewed as based on "private information."[4]

One further characteristic distinguishes clearinghouses processing contracts for immediate performance from those entailing future performance. This is that once settled, the obligations of all parties are complete: Neither buyer or seller has any further obligations to each other or to the clearinghouse, nor does the clearinghouse have further obligations to the buyer or the seller. As detailed in the next subsection, this differs from the requirements of clearinghouses processing contracts for deferred performance, where settlements generally require periodic performance by one or both parties.

Contracts for Deferred Performance

In the case of executory contracts—the general legal term for forwards, futures, or options—counterparty defaults are especially important because these contracts defer the performance of a significant portion of their obligations. For example, forward and futures contracts obligate their holders to deliver either cash or the underlying at a future date; options require those writing the option to buy (for a put) or sell (for a call) the underlying during some time interval. First we provide a general description of the clearing operations associated with futures exchanges, then, using futures as a point of departure, we describe clearing for exchange-traded options.

The performance risk of executory contracts is crucial to an understanding of the special requirements of their clearing and settlement. To see this, consider a forward contract between two parties, in which the "short" promises to deliver a specified good on a specified date (the delivery date) to the "long" at a price specified on the date of the contract. As is usual, the contract requires no payment until the contract's expiration date. Although both parties enter into the contract expecting to benefit from transacting, circumstances can subsequently arise that make fulfilling their respective commitments at the expiration date excessively burdensome. In those instances, recourse to limited liability and/or the cost of legal enforcement enables losing counterparties to opt out of completing commitments (i.e., defaulting) in certain instances.[5]

The motivation for default often comes from a change in the price of the delivered item. As this price increases, the short has an incentive to renege on its commitment to sell at its contracted price, which has developed to be a below-market price (e.g., if the short owns the good, by fulfilling the delivery commitment it forgoes the opportunity to sell to a third party at a higher price). Similarly, as price falls, the long has an incentive to renege on its obligation to buy the good at its contracted price. A long (or short) wanting to lock in the price it will pay (receive) at contract expiry needs to be concerned not only with the prices it is offered; it must also factor in the likelihood that a given counterparty may default should prices move unfavorably. Applying the well-recognized logic that as time increases more things can happen, we can conclude that the greater a contract's "futurity"—the time remaining until performance under the contract is required—the greater is the likelihood of realizing a change in price sufficient to induce default. Thus, a party wishing to "lock in" a price at a future date needs to be concerned about the financial integrity of its counterparty.

To be sure, the default risk associated with executory contracts is present to some extent for contracts requiring immediate performance. However, the futurity of executory contracts means that both the extent of loss and its probability are substantially higher than for contracts requiring immediate performance. This, in turn, means that clearing executory contracts requires undertaking extra steps to mitigate these risks. Specifically, clearing for executory contracts consists of systems for intermediating trades so as to lessen losses from counterparty defaults. This is accomplished by imposing standards that mitigate risk exposures. These standards enable intermediaries to commit in advance to accepting traded contracts for clearance and settlement. Hence, if a party to a trade is not a member of the clearinghouse (*clearing member*), the party establishes a contractual relationship with an intermediary (the clearing member), who is responsible for fulfilling the contract (e.g., delivering the goods or cash promised to the trader). In this way, intermediaries, which are high-credit-quality firms, are interposed as counterparties for trades made by exchange members that are not clearinghouse members, essentially taking on positions established by the initial buy and sell sides of the trade.[6] That is, the promises made to the long and short positions in the initial trade become the responsibility of clearing members. In turn, the clearinghouse stands as an intermediary between the clearing members on the two sides of the trade, again assuring the promises made to the long and short sides of the trade. The strength of the arrangement then depends on the financial strength of the clearing organization.

Exhibit 19.1 illustrates the arrangement. The top panel lists initial positions that will be used as a point of departure for the three remaining panels. To simplify the exposition, the members listed as *Long 10, Short 7,* and *Short 3* are members trading for their own accounts. Member names also relate their positions so that *Long 10* has 10 contracts for deferred purchased at the price 98 and *Short 7* and *Short 3* have, respectively, 7 and 3 contracts to sell at that price. In the example, we assume that these traders have no other positions. These trades are intermediated by the clearing arrangement, which consists of three clearing firms: *A, B,* and *C.* Both *Long 10* and *Short 3* will contract through *B* and *Short 7* will contract through *A.* In the example, *C* will not clear any trades.

Case I relates how the arrangement works absent default. In this hypothetical, the price has risen to 101 on the expiration date, an increase of \$3. This means that *Long 10* has gained \$30, while the shorts have lost \$21 and \$9 respectively. Specifically, *Short 7* is required to pay the clearing member *A* \$21, and clearing member *A* is required to pay the clearinghouse that same amount. At the same time, *B* is now entitled to a payment of \$21 from the clearinghouse, and that same

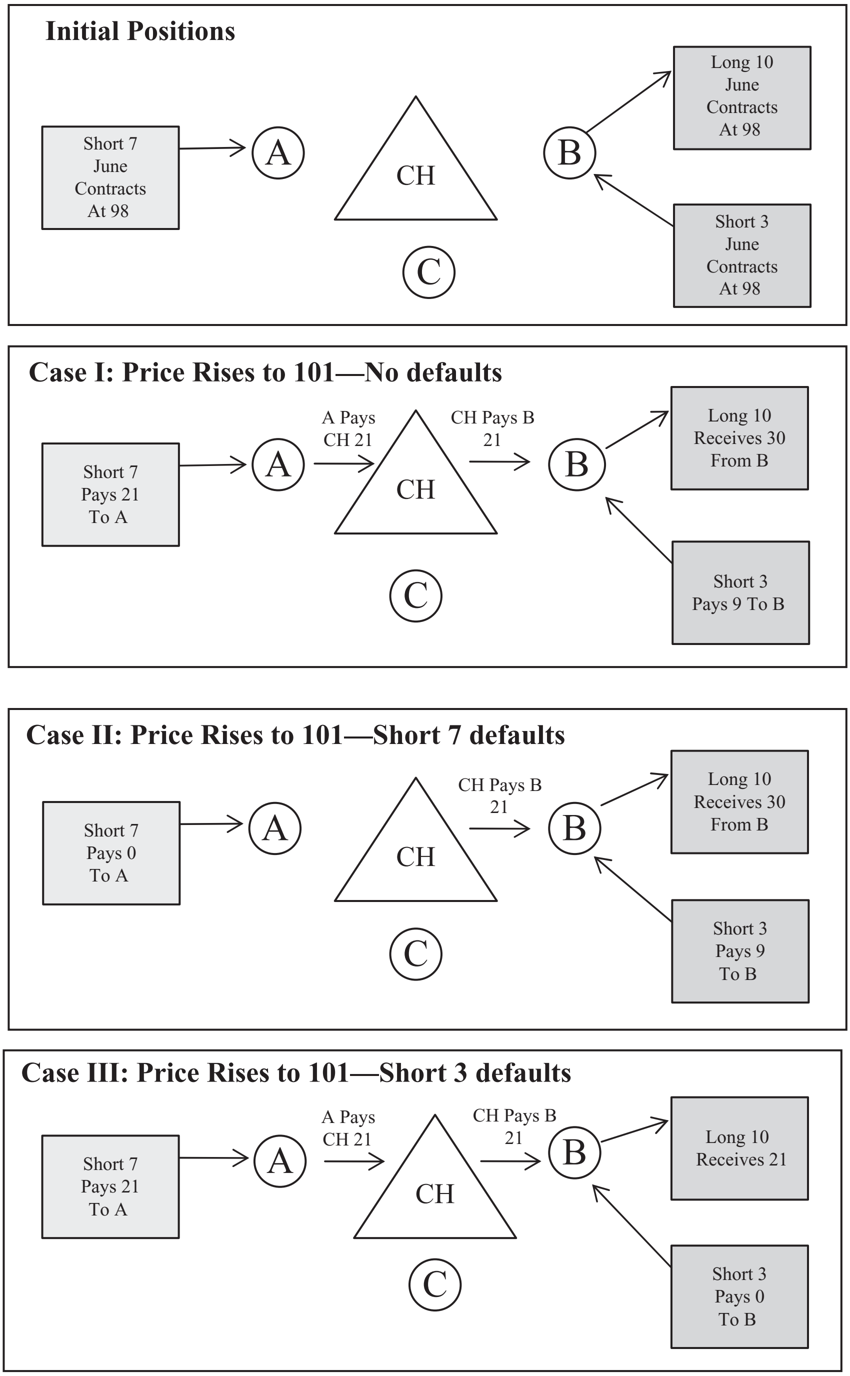


Exhibit 19.1 Flow of Funds Between Parties in Clearing Arrangement

amount is to be paid by *B* to *Long 10.* In addition, *B* will collect $9 from *Short 7,* which he also pays to *Long 10.*

The remaining two panels depict two different default scenarios. Case II depicts the multiple layers of intermediation employed at the clearinghouse. Suppose that *Short 7* defaults on its obligation to pay $21 to clearing member *A.* That nonpayment does not eliminate clearing member *A*'s liability to the clearinghouse of $21. Similarly, the clearinghouse remains liable for a payment of $21 to clearing member *B,* who is liable for a payment of $30 (including the $9 *Short* 3 paid him) to *Long 10.* Should *A* also prove unable to make that payment, the clearinghouse, as the intermediary between clearing members, remains liable for the amount due to *B,* with *B* then passing that amount to *Long 10.*[7] In this way, *Long 10* receives payment, even though both *Short 7* and his clearing member defaulted.

It is important to see that *the clearinghouse guarantee is between clearing members*; that is to say, once clearing members are substituted for the traders who established the position, the clearinghouse becomes the buyer to every seller and seller to every buyer. The distinction is important in that the clearinghouse does not become the buyer of the positions of *Short 7* and *Short 3,* nor the seller to *Long 10.* It is instead the buyer of 7 contracts cleared through *A* (originated by *Short 7*) and the seller for 7 contracts cleared through *B* (originated by *Long 10*).

To illustrate the significance of this distinction, Case III examines what happens when *Short 3* fails and its clearing member *B* is unable to cover the loss. When the *Short 3* due-from amount fails to arrive, clearing member *B* becomes liable for the defaulted amount. If clearing member B is unable to make that payment, *Long 10* will simply not receive the entire $30 payment due; the clearinghouse has no direct responsibility to *Long 10.*

Case III illustrates that even when the clearinghouse is solvent, each trader's guarantee of receiving the payment implied by the price change (i.e., *Long 10* receiving $30), is not absolute. In that example, the cause of *Long 10* not receiving payment was that clearing member *B* had insufficient resources to cover *Short 3*'s default; *Long 10* does not benefit from the clearinghouse guarantee concerning the amount due from Short *3.*[8]

The examples illustrate two important features of the clearinghouse system.

1. The clearinghouse system allows contract counterparties to receive the payments they are due, even when those required to pay are in default.
2. This system does not always result in counterparties receiving the full payments associated with the price change, even if the clearinghouse is solvent (as Case III illustrated).

These examples also illustrate how clearing standardizes default risk across potential trading partners. Whether *Long 10* receives his full credit of $30 depends on the creditworthiness of those traders (and their clearing members and the clearinghouse) who have outstanding short positions at the time the price changes from $98 to $101. It does not depend on who *Long 10* traded with to acquire his position. For example, *Short 3* may have acquired his position through a trade with *C. C* in turn could have been unwinding the short position he had acquired through an earlier trade with *Long 10.* However, the risk that *C* defaults becomes irrelevant to his trading partners once *C* acquires a zero net position. Hence, a key consequence of clearing is that it standardizes counterparty risk by making default risk independent of trading counterparty.

It is worth noting that while clearinghouses have no legal obligation to make nonmembers whole when a member defaults (as in Case III), as a practical matter, a clearinghouse may choose to make such payments to nonmembers. If traders like *Long 10* believe that when their clearing member defaults, it will lead to their not receiving full payment, they will be less willing to trade on the exchange. To enhance the attractiveness of its trading platform, an exchange and its associated clearinghouse may be willing to make payments that they are not legally obligated to make in order to convince traders of the financial integrity of the clearing process, and thereby preserve the ongoing value of the clearinghouse. That is, if the resulting increase in the ongoing value of continuing operations exceeds the cost of making such payments, then it can be in the clearinghouse's interest to make such traders whole.

This insight offers a route for thinking about questions of clearinghouse ownership structure. Structures providing greater going-concern value to the owners of the clearinghouse should be more likely to undertake the duties of defaulting members. This point has an important policy aspect in that clearinghouse failures often are regarded as systemic in nature. A clearinghouse whose members have strong incentives to head off losses that may result in financial system-wide problems should be preferred over a clearinghouse less inclined to take those steps.

Enhancing the credit-worthiness for long-lived contracts like futures is costly to clearinghouses, prompting them to employ a variety of rules to manage those costs. The rules by which clearinghouses reduce the costs of assuring performance fall into two general categories: rules mitigating the likelihood of counterparty default and rules creating funds for the clearinghouse to draw upon when defaults do occur.

The principal means by which clearing systems mitigate the likelihood of default is through mark-to-market adjustments. Marking to market refers to periodic determination of the gains and losses incurred by every position since the previous mark. This is accompanied by collection against losses and distribution of gains. Generally, marks are set at the end of each trading session, based on the then-current market price.[9] Hence, for position holders with extant short positions, a price increase results in the clearinghouse collecting funds from them to cover losses. Those funds are then distributed to the owners of long positions. The reverse occurs when a price declines. The effect is that all open contracts are essentially replaced by contracts established at the most recent price.[10] These flows of money are called variation payments and are paid into variation margin accounts. Hence, the kinds of flows described in Exhibit 19.1 occur on a daily basis and are examples of variation margin payments.

Certain margining arrangements augment the mark-to-market system. When positions are initially established, contract sides are required to post a performance bond, known as the initial margin.[11] Typical initial margin amounts range from 5 to 20 percent of the notional value for futures trades.[12] Subsequently, as the market price changes and some variation accounts become negative, the initial margin serves as an auxiliary account that can be drawn on to cover, temporarily, any deficit in the variation account. If the value of the initial margin account falls below a fixed level, known as the maintenance margin level, the position holder is required to bring the margin account balance to its full level. Failure to meet this requirement results in liquidation of the position. For customer accounts, the customer's position is effectively taken over by the clearing member. The combination of mark-to-market and performance bonds means that the maximum loss to the clearing member will be equal to the single-period price change, minus the amount in the maintenance margin account (prior to marking to market), minus the member's costs of selling off the position. Similarly, should a clearing member fail to meet its margin requirements, the clearinghouse's losses are limited to the single-period price change, net of the amount in the margin account and the costs of selling off the position.

Of course, the size of the margin also affects the likelihood of default. To illustrate, if margin levels are chosen so that 95 percent of one-day price changes are less than the initial margin balance, then the loss will exceed the initial margin roughly once in 20 trading days. As a practical matter, incidences of default occur much less often than 5 percent of trading days. This is because the holders of open positions often have other positions with positive margins (which the clearinghouse can appropriate when default occurs on one position), or they opt to replenish their initial margin accounts so as to retain their positions. However, when defaults do occur, the available margin balance lessens the amount of loss. In this way, these two provisions serve to reduce the clearinghouse's cost of assuring execution.

Exhibit 19.2 presents some examples of initial and maintenance margins for some of the large volume products traded on U.S. futures exchanges as of October 18, 2007. A comparison of the two final columns indicates that the maintenance margin is typically about the same size as the largest one-day price movement over the previous year.

How Margin Works

To see how margins protect the clearinghouse, consider a change in the price of soybeans. Suppose the price agreed to by two parties on a 5,000-bushel soybean futures contract traded on the Chicago Board of Trade (CBOT) is $10.00 per bushel. As noted in Exhibit 19.2, the initial margin required of contract holders is $2,700. Further, suppose the settlement price for soybeans on trades made on the following day is $10.40. In that case, the short's paid-in margin is reduced by $.40 per bushel, or a total of $2,000. This means that the short only has a remaining margin of $700, and is required to make a payment of $1,300 to satisfy the maintenance margin requirement of $2,000. If one side fails to make the requisite payment (defaults), he or she forfeits the remainder of the margin that had previously been paid, and his or her position is taken over by the party holding the margin (e.g., the clearinghouse).

These required margin payments play an important role in reducing the clearinghouse's cost of assuring contractual performance. As described, they provide clearinghouses with funds to meet payment obligations when defaults do occur. More important, the requirement of a large maintenance margin weeds out

Exhibit 19.2 Margin Levels of Large Volume Contracts and Largest Price Movements over a One-year Period Ending 10/18/2007

Futures Contract	Initial Margin[a]	Maintenance Margin	Largest One-Day Price Movement, Previous Year[b]
Chicago Board of Trade			
10-year Treasury note	$1,148	$ 850	$1,047
5-Year Treasury note	$ 878	$ 650	$ 766
Soybeans	$2,700	$2,000	$2,500
Wheat	$2,025	$1,500	$1,500
Dow Jones Industrial Average ($5)	$3,125	$2,500	$2,370
Chicago Mercantile Exchange			
3-month Eurodollars	$ 550	$ 550	$ 318
E-Mini S&P 500 index	$3,150	$3,150	$2,865
E-Mini Nasdaq 100 index	$2,600	$2,600	$1,745
Japanese yen	$2,000	$2,000	$3,575
Euro FX	$1,500	$1,500	$2,013
NY Mercantile Exchange			
WTI crude oil, sweet, light	$3,300	$3,300	$3,420
Natural gas	$7,500	$7,500	$9,700
Gold	$2,500	$2,500	$2,310
Silver	$3,000	$3,000	$5,265

Notes:
[a]Initial margins differ for members and nonmembers. The listed margins are those for exchange members.
[b]Soybeans and wheat are subject to maximum one-day price movement limits. Hence, the maximum change reported in the table may understate the maximum one-day change in the underlying value of the commodity.

counterparties whose positions become sufficiently negative that they are likely to choose not to fulfill their obligations under the contract (e.g., longs that will not pay the contract price at expiry). Without the necessity of making variation margin payments, counterparties that experience losses in their positions could retain those positions until contract expiry in the hope that market prices would move favorably. This amounts to their having a free option; if the price moves in a favorable direction, they are better off, while movements in the opposite direction would incur no further loss (since the account would be in default in any case).

The mark-to-market process, along with the requirement of paying variation margin, essentially converts the long-term contract into a series of daily contracts spanning the same interval. Upon recognizing that a counterparty has insufficient resources to fulfill one of these daily settlements, the intermediary can liquidate the position rather than, in effect, granting an option paying off only if the prior price change is reversed.

Option markets differ in an important respect from futures markets. Whereas in forward or futures markets both counterparties are obligated to fulfill their sides of an executory contract, in options markets only the sell side has that obligation. The buy side of an option contract has instead a right to act and presumably will rationally exercise that right. Hence, measures taken against default of options contracts focus on the sell side. Specifically, the seller of the option must deliver cash, securities, or futures contracts should the option buyer exercise its rights under the contract. This obligation means that the seller of an uncovered put[13] (call) is in a similar position to the long (short) side of a futures contract. If price moves adversely for the trader (e.g., for the seller of a call, when price rises), the potential loss to the seller of the option is limited only by the agent's wealth. Not surprisingly, the same basic requirements that are made of traders in futures markets are made of sellers in options markets. That is, the seller of a call or a put must make an initial margin payment and then make variation margin payments if the underlying contract price moves against the trader.

For both futures and options, initial margin is required of all parties in the processing chain, and these must not fall below specified maintenance levels. That is, if the trader is not a clearing member, then the clearinghouse requires both that the trader post margin with the clearing member and that the clearing member post margin with the clearinghouse. In this process, the clearing member becomes the legal counterparty to the nonclearing member, and the clearinghouse becomes the legal counterparty to the clearing member when the trade is cleared. For a trader who is not an exchange member, an additional step may be required. A nonmember trader must use the services of a member who is a broker (or Futures Commission Merchant—*FCM*—in the case of futures) to make the trade. The broker/FCM will charge the trader a fee for that service and also will require initial margin and variation account payments (which are again subject to a clearinghouse-set minimum) from the trader.[14]

This system of successive obligations provides position holders a second means of protection against default. As discussed in connection with Exhibit 19.1, should a trader default, the next agent in the processing chain is required to cover the default. For example, if a retail trader does not make his or her variation margin payment, the FCM/broker who executed the trade is still required to make a variation margin payment to the clearinghouse (or the clearing member if the broker/FCM is not a clearing member). Clearing members generally are required to set aside funds to cover such customer defaults (e.g., the capital charge) and are required to meet capital adequacy standards.[15] Clearing members are audited by clearinghouses to ensure that they remain in compliance with these standards.

Finally, as illustrated in Exhibit 19.1, should a clearing member default on his variation margin payment to another clearing member, the clearinghouse is responsible for paying the clearing member who is due a payment. If the trader who is due the payment is not a clearing member, then the clearing member with whom he cleared is responsible for paying off the trader, once the clearing member gets paid. Clearinghouses have two kinds of devices to fulfill their obligations when clearing member failure occurs.[16]

1. Clearinghouses accumulate reserves (including required contributions from clearing members) to cover the losses associated with clearing member default.
2. Agreements among the clearing members establish how losses beyond the level of reserves will be allocated among the nondefaulting clearing members. All clearinghouses mutualize loss exposure among the clearing members, although the exact mechanism used varies across exchanges.[17]

Through these devices, the clearinghouse provides traders with a high level of confidence that the gains from any trade executed on the exchange eventually will be realized by the trader. In practice, no U.S. clearinghouse has failed to settle the outstanding positions of any clearing member. Some examples of non-U.S. clearinghouses that failed to settle customer positions are provided in Hills, Rule, Parkinson, and Young (1999).

The preceding discussion implies that clearinghouses act to protect themselves from losses by employing a variety of tools. Their loss-avoidance efforts provide assurances that the obligations entailed in long-futurity contracts will be met, specifically that both contract sides can expect to realize the payments or deliveries to which they are due. This results in a lessening of both the deadweight losses from collecting against defaulted contracts and the costs associated with replacing positions.

We would anticipate that the tools are substitutes for one another. In fact, in the United States, the clearinghouse for equity options uses a substantially different combination of these tools than do the clearinghouses for futures and options on futures. For equity options, which are all cleared by the Options Clearing Corporation (OCC), initial customer margins for the party selling the option are set by the Federal Reserve, currently at 20 percent of the price of the underlying equity plus the premium paid by the option owner, which is higher than the margin set by the futures clearinghouses. Because the level of assurance provided by margin alone is higher for equity options than futures, other things being equal, the same level of assurance can be provided by the equity options clearinghouse with lower reserves. Evidencing this point is the fact that futures and futures options clearinghouses require substantially higher levels of member reserves and guarantees than the equity options clearinghouse; the OCC requires members to contribute $150,000 to the clearinghouse reserves, while clearing members are required to contribute a minimum of $500,000 to the Chicago Mercantile Exchange (CME) clearinghouse.

CLEARING AND LIQUIDITY

Traders gain two key advantages from trading on an exchange rather than over the counter. The first advantage, which was discussed earlier, is that counterparty default risk is mitigated by the presence of the clearinghouse. A second advantage is that trading is likely to be more liquid on the exchange than elsewhere. Liquidity is the ability to transact immediately with minimal price adjustment; that is, in a liquid market, traders expect that most orders can be filled with minimal effect on prices and can be unwound at a low cost (i.e., low bid-ask spreads).

The ability of an exchange to provide a liquid market depends on the presence of considerable heterogeneity among traders on that exchange. If all traders have similar initial positions (e.g., all are corn farmers wishing to hedge the risk of price changes), they will find it expensive to trade. When all hedgers have the same cash-market exposure and are risk averse (which is why they are hedging in the first place), they all would like to trade in the same direction (e.g., corn farmers wishing to undertake short positions). In this case, each trader taking the opposite side of a contract with a hedger will have to be paid a large premium to take the opposite side of the hedging contracts. Hence, markets in which all traders have the same initial positions tend to be characterized by high bid-ask spreads. In contrast, the heterogeneous mix of trading interests needed to obtain higher liquidity leads to larger exchange memberships.[18] This implies a lower likelihood of knowing the creditworthiness of one's counterparty; from another perspective, knowing one's counterparty requires that exchange participants undertake greater outlays for credit monitoring.

As emphasized, the fact that trades on an exchange are cleared means that the creditworthiness of the counterparty in a trade is irrelevant to a trader. This can be viewed as one additional kind of standardization, just like the standardization of other features of the transaction, such as product characteristics, timing of payments, and the like. What this means is that from an individual trader's perspective, the only relevant difference between potential trading partners in a cleared futures contract is the price they might offer. In comparison to forward contracts, futures contracts allow traders to have good information about the default risk they face without acquiring information about the counterparty in their trade.

These advantages of clearing help explain why organized futures exchanges represent an important venue for trading. How futures trading in an instrument is structured (e.g., the extent of interexchange competition) can be influenced by how clearing is organized. Two features of futures trading tend to lead to concentration of trading to a single venue. The first feature is what is frequently called the liquidity externality.[19] The idea is that trading in a given contract tends to congregate in a single venue because individual traders attempt to minimize their trading costs (including the impact of a transaction on the market price and the expected search costs). Individuals accomplish this by searching across potential counterparties at the location that they think will be the most likely one for them to find a suitable transaction. As success at achieving low trading costs becomes the norm at a particular venue, transactions become increasingly costly at alternative venues. In futures markets, this is referred to as the first-mover advantage in that most often the first exchange to achieve a liquid contract market remains the dominant venue for trading that contract.[20]

The tendency for trading to congregate in a single venue due to the liquidity externality is reinforced by a second advantage that larger venues possess. Traders value the ability to unwind their positions (i.e., return to a zero net position from their initial long or short position).[21] Because the trader's contract is with a clearing member and ultimately with the clearinghouse, a position in a futures contract can simply be "offset" by the trader making an opposing trade—going long against a previously held short position or short against a previously held long position—on the same exchange. In fact, it is widely recognized that most futures market positions are closed through offset rather than held until expiration.[22] Traders prefer to trade instruments for which the costs of unwinding are low (e.g., in contracts that will be liquid in the near future).

At this point, it is worth distinguishing between an instrument and the venue(s) on which it trades. Some instruments, such as equities and equity options, are traded on multiple exchanges, which means that a trader can make a trade on one exchange and unwind it on another. Positions in other instruments, such as most futures and futures options, can be unwound only on the exchange where the initial trade was made, even if other exchanges trade quite similar instruments (such as the gold futures contracts on the New York Mercantile Exchange and the

CBOT in 2005 to 2008). In these cases, the instrument is synonymous with the exchange. Trading on smaller exchanges will be more attractive for contracts for which a trade initiated on a smaller exchange can be unwound on the larger one. Conversely, when each contract is traded on only one exchange, a trader who is considering taking a position on a small exchange needs to be concerned not only about the costs of establishing a position on the smaller market but also the costs of closing the position (since closing must occur on the less-liquid exchange). This tends to make it more difficult for the smaller exchange to overcome the liquidity externality and thus tends to reduce interexchange competition.

As detailed later in the chapter, the rules regarding interexchange clearing arrangements affect the cost of establishing a position on one exchange and unwinding it on another. As such, these rules can affect the viability of smaller exchanges. Some suggestive evidence for this premise can be seen by comparing equity trading and futures trading. In the United States, equity trades are cleared by the Depository Trust Clearing Corporation (DTCC), regardless of the venue on which the trade takes place. Similarly, all equity options are cleared by the Options Clearing Corporation (OCC). In contrast, trades made in futures and futures options are cleared by exchange-designated clearinghouses. Importantly, each futures exchange chooses whether contracts opened with trades on other exchanges can be offset with trades at that exchange.[23] As will be discussed, this seems to have influenced interexchange competition for these kinds of instruments.

COMPETITION BETWEEN EXCHANGES

Whether because of the differences in clearing or other factors, the nature of competition between exchanges differs between futures and futures options on one hand and equity and equity options on the other. Specifically, for equity and equity options, the same instrument is actively traded on multiple venues. For example, options on Microsoft equity are actively traded on all six U.S. options exchanges, and a substantial percentage of that trading occurs on each of several exchanges. In contrast, for virtually all products in the United States, trading on individual futures and futures options products tends to be concentrated on a single exchange.

Based on the premise that interexchange competition leads to lower trading costs, some commentators have advocated mandating changes in the clearing arrangements for futures markets that will promote direct competition on individual futures contracts across exchanges.[24] Those advocating change generally suggest that policies be implemented that result in enhanced offset between trades made on different exchanges in a given contract. For example, Harris (2006) suggests separating ownership of clearing facilities from ownership of the exchanges, as is the case with equity options. The idea behind this is that clearinghouse owners would then generally seek to promote competition between exchanges in order to increase demand for clearing the contract. John Damgard, president of the Futures Industry Association (an organization of FCMs), suggests rules mandating that exchanges allow clearing choice, whereby any clearinghouse approved by the Commodity Futures Trading Commission would be allowed to clear any contract.

While none of these proposals for promoting competition on individual contracts has been specified in detail, certain issues will almost certainly be relevant to any such plan.

Nature of the Clearing Organization

A common component of the suggested changes to futures markets is that a trade initially made on one exchange, with one clearinghouse, can be fully offset by a trade made on a different exchange and/or by using a different clearinghouse. While the means by which this might happen has not been specified, we can envision two kinds of mechanisms: a central clearinghouse or competing clearinghouses.

Central Clearinghouse

As noted, for both equity and equity options trading, all trades are cleared by a central clearinghouse, regardless of its trading venue. This structure allows offset between trades made on different venues. While different in many ways from the clearing structure used for futures and futures options, the two structures do have an important common element. In both cases, contracts between clearing members are replaced by contracts between clearing members and the clearinghouse, so that the clearinghouse has a zero net position in those contracts; that is, absent default by a clearing member, the clearinghouse is neither net long nor net short in any contract obligation. This reduces the clearinghouse's risk and increases its ability to assure performance of contractual obligations.

One benefit of having a central clearinghouse is that it mitigates the advantage the largest clearinghouse has in unwinding positions and encourages competition between exchanges. The presumption is that such competition will reduce trading costs, and, in fact, the evidence from financial markets is consistent with this. For example, competition between exchanges has been shown to reduce bid-ask spreads on equity options (e.g., DeFontnouvelle, Fishe, and Harris [2003]).

Trading costs consist of the effective bid-ask spread a trader pays, plus exchange transaction fees and clearing fees. While exchange competition resulting from a central clearinghouse seems likely to reduce the first two components, the effect of a central clearinghouse would seem to raise clearing fees relative to (imperfect) competition between integrated exchanges/clearinghouses. This is because any reduction in the first two components due to the creation of the central clearinghouse will increase the demand for clearing services, which would generally induce the monopoly clearinghouse to raise its fees. Therefore, the net effect of moving to the central clearinghouse might rather lead to an increase in trading costs.[25]

Moving to a central clearinghouse model will reduce trading costs if the clearinghouse is somehow prevented from using its monopoly position. In practice, both the DTCC and the OCC are nonprofit corporations owned by their clearing members. The nonprofit status means that the clearinghouses are constrained in their ability to distribute revenues to their owners. Both are often described as utilities, suggesting that their pricing of clearing services is constrained to approximate the average cost they face in providing the clearing service.

The nonprofit status, however, would not by itself eliminate the incentive to use their monopoly positions to increase member profits (Pirrong, 2000). For example, an indirect means of providing income to its owners would be for a clearinghouse to set a large difference between the margin it requires of clearing members and the minimum margin clearing members charge other members and nonmembers.[26]

Hence, the fact that the DTCC and the OCC are nonprofits does not mean that total trading costs are at "competitive" levels. That is, the goal of reducing trading costs by switching to a central clearinghouse is not necessarily achieved by assuring that the central clearinghouse earns zero profits. This goal may require more intrusive reviews of clearinghouse practices.[27]

Competing Clearinghouses

An alternative to the central clearinghouse model is a model in which multiple clearinghouses could clear the same contract, which is sometimes referred to as clearing choice.[28] As is done today, exchange members would contract with a clearing member of one of these clearing organizations. Differing from current practice, the two sides of any trade may be cleared through separate clearinghouses. Since this model of clearing has not been adopted, many practical considerations have not been addressed. Most important, the precise nature of relationships between the clearinghouses is yet to be established. This is important because of the nature of a clearinghouse's net position in this model. In contrast to the current model and the central clearinghouse model just described, individual clearinghouses generally will not have a zero net position at the end of the trading day. To see this, consider a clearing member whose sole affiliation is with clearinghouse A. Then all of the customers who cleared their trades through that member will ultimately have a position with clearinghouse A. There is no reason to anticipate that the contracts of all members of any one clearinghouse will net to zero. Thus, clearing choice generally will result in clearinghouses becoming net long or net short in each contract. This makes clearinghouses more vulnerable to subsequent market volatility.

We envision three possible solutions to this problem. One solution would be for each clearinghouse to retain the nonzero net positions and address its risk exposures by a combination of greater clearinghouse reserves and/or lower execution assurance (i.e., a greater likelihood of clearinghouse failure). Both higher reserves and lower execution assurance are costly to market participants.

A second solution would be for the two clearinghouses to voluntarily work together to mitigate the risk each faces from nonzero positions (e.g., by offsetting with each other). However, one clearinghouse may have a strategic interest in leaving another clearinghouse with increased risk. For example, if there are two clearinghouses for a contract, both have the same size net position in terms of the number of contracts, even if one has a much larger share of the total number of contracts. The net position will constitute a greater risk as a percentage of clearing revenues for the small clearinghouse and will represent a more significant cost for the small clearinghouse. As such, it may be in the large clearinghouse's interest to retain its nonzero net position in order to disadvantage its rival.

A third solution would be to have a central "clearinghouse of clearinghouses." That is, this clearinghouse of clearinghouses would play a role similar to that taken by the central clearinghouse in the model described earlier. The central clearinghouse would become the counterparty to all of the other clearinghouses and hence attain a zero net position. Of course, if a central clearinghouse is the ultimate clearinghouse for other clearinghouses, then one has to question the additional value of the competing clearinghouses. That is, why introduce an additional transaction? It would be simpler to use the central clearinghouse for all clearing rather than adding a second step in the clearing process. Whether clearing choice exists in addition to this central clearinghouse or not, the issue raised earlier remains: Given the monopoly position of the central clearinghouse, what prevents the central clearinghouse from exercising its monopoly position to increase transaction prices?

Innovation and Clearing Structure

The goal of those proposing alternative clearing arrangements is to reduce trading costs, essentially by reducing an exchange's ability to set prices for services above their costs. High prices might take the form of high transactions fees or rules that lead to higher bid-ask spreads. In general, restraining a firm's ability to price above cost is efficient in a static sense; setting price closer to marginal cost generally leads to greater gains from trade. In a dynamic sense, however, the welfare consequences of restraining monopoly power are less clear-cut. The potential for pricing above cost often provides the impetus for innovation. There is a long literature on the welfare trade-off involved in setting intellectual property rules that outlines the basic issues in the trade-off between static and dynamic efficiencies.[29] Of course, this trade-off is also implicit in the intellectual property policies of governments (e.g., 17-year patent protection for significant innovations).

Thus, to the extent that innovation is important in financial markets, allowing some pricing above cost may be socially valuable. The importance of innovation by financial clearinghouses plausibly varies across instruments. Securities are created by the firms that issue them; the role of financial markets is to provide a low-cost, transparent environment in which to trade them. In contrast, futures and futures options are created by exchanges. There is considerable risk in creating new futures contracts. Not only are there significant costs borne by exchanges in developing new contracts, but in addition, most futures contracts are unsuccessful.[30] Equity options occupy a middle ground between futures and equity in terms of the costs of development.

This suggests that, even if a central clearinghouse that acts like a utility is an optimal means of organizing an industry for equity and equity options, it may not be for futures and futures options. This is because encouraging innovation may be more valuable for futures than for equities.

CONCLUSION

While clearing is a behind-the-scenes aspect of exchange operations, it is crucial to the workings of an exchange. In addition to the standardization of contracts with respect to their required deliveries and payments that exchanges achieve, clearing standardizes credit risk. This arrangement completes the commoditization of the terms of trade, in that each contract is for a standard product; but in addition, the financial backing of every trade is also standardized. Hence, traders need not negotiate over what is being traded, nor need they consider the inescapable heterogeneity in financial ability of their trading counterparts.

The nature of the relationship between an exchange and its clearinghouse can have a significant effect on trading costs. The common clearing model used for equities and equity options seems to promote competition between exchanges, which in turn appears to lead to lower bid-ask spreads. Whether this model is appropriate for futures markets will likely be a topic for debate over the next few years.

ENDNOTES

1. In recent years, similar arrangements have become more common for OTC trades. Specifically, creditworthy institutions known as derivatives product companies (DPC) have arisen that take one side of an OTC trade, allowing the trader to have a high level of assurance against counterparty risk. The DPC in turn achieves its high credit rating by trading away its net market risk (typically to an affiliate), limiting its counterparties to creditworthy traders, requiring margin deposits by those counterparties, and reducing risk through futures trades. See, e.g., Kroszner (1999). See Moser (2000) for an economic history of the evolution of clearing systems used by futures exchanges.
2. Immediate performance is somewhat of a misnomer. More likely the transaction will be effected no earlier than the next day, and in many instances several days will pass before the transaction can be said to be effected.
3. So-called straight-through processing can also be compared to a computer program in that all of the necessary steps are identified in advance of the trade. On execution of a trade, those steps are executed immediately.
4. The other side of this coin is that traders having information can trade anonymously on that information. This enables insider trading, which possibility provides a rationale for securities laws requiring insiders to disclose their trading activity.
5. Kane (1980) refers to the ability to opt out as a nonperformance option. His premise is that trading is valuable. Were trading to be limited to only those cases where counterparties were certain to perform, then trading activity would be substantially circumscribed. Allowing nonperformance options is efficient if the loss associated with limiting trade to circumstances without nonperformance options is sufficiently large.
6. There can be additional steps in this process if the initial trader is not a member of the exchange. In such a case, the trader uses the services of an exchange member, who serves as a broker. If the broker is not a clearing member, then the broker stands as an intermediary between the clearing member and the trader.
7. When A defaults on its required payment to the clearinghouse, its status as a clearing member is terminated by the clearinghouse. Subsequent payments from B or C due to A for members formerly clearing through it will be rerouted to avoid misuse of those funds by the failing clearing member A.
8. As a practical matter, nonpayment by a clearing member is rare, in part because the clearinghouse has incentives to ensure the financial solvency of clearing members.
9. The price used for daily marking (or settlement) typically is based on trades made in the last portion of the trading period (e.g., the last 2 minutes). In special cases—most often for thinly traded contracts—a committee may determine that these trades do not properly reflect market circumstances. In those instances, the committee will specify the settlement price based on additional factors.
10. The legal term for this is novation, the replacement of one obligation with another as agreed upon by the parties to the contract.
11. This discussion has focused on simple positions in which the trader's portfolio consists solely of a single position in a single contract. When a trader has multiple positions in the same contracts, such as spread positions (i.e., a long position in a contract expiry in a specific month and a short position in the same contract with a different month's expiry), margin requirements are different. See Rudd and Schroeder (1982) for a discussion of margin calculations for more complex portfolios.
12. The notional value is the product of the negotiated price of the futures contract (e.g., \$10/ bushel for soybeans) and the quantity of items specified by the contract (e.g., 5,000

bushels of soybeans). Although initial margin can range up to 20 percent of the notional value, margins closer to the 5 percent figure are more typical. See Baer, France, and Moser (2005).

13. An uncovered option is one in which the seller has no position on the underlying instrument. As explained in note 11, a portfolio consisting of positions in related instruments—such as writing a call option and owning the underlying security (known as a covered call)—has different rules for setting margins.
14. This is true of electronic trades as well as traditional floor trades. With an electronic trade, the nonmember trader may enter his or her trade directly into the system, but only through a broker/FCM's facilities.
15. For example, FCMs are required to have adjusted net capital equal to at least 4 percent of the aggregate value of the customer margin accounts. A 1986 National Futures Association survey found that virtually all FCMs had at least twice that ratio, and the majority had greater adjusted net capital than the value of the customer margin accounts.
16. For example, see the description of CME's financial safeguards, available at http://www.cmegroup.com/company/membership/types-of-membership.html? show=Clearing
17. Specifically, some clearinghouses have contracts with insurance companies for covering losses beyond the paid-in reserves, while other clearinghouses can legally assess clearing members for additional funds. Clearing members accept this financial responsibility in exchange for the right to trade directly with the clearinghouse, which allows them to have lower trading costs than nonclearing members.
18. This point about risk premiums can be seen by considering a futures markets whose participants are all hedgers. One-half are long in futures and are seeking to hedge future cash purchases (e.g., flour mills), the other half are short hedgers seeking to hedge future cash sales (e.g., grain farmers). Because each side is willing to pay the other to reduce their respective risk exposures, the costs offset and the necessary premium is smaller.
19. See, e.g., Hendershott and Jones (2005).
20. What is important to note is that competition between exchanges in a single contract tends to be transitory, as all but one exchange drops the contract after a short competitive period. While there are instances in which the exchange that introduces the contract does not prevail in this competition (the Bund contract being the preeminent example), Holder, Tomas, and Webb (1999) find that which exchange introduced the contract is an important factor in determining the outcome of that contest.
21. As Edwards (1983) notes, clearing greatly reduces the cost of unwinding because it enables a contract holders to offset their position through a trader with any counterparty rather than negotiating with their initial counterparty (as one would do for a forward contract).
22. According to the Chicago Mercantile Exchange (2006), only about 3 percent of futures contracts result in physical delivery.
23. In some cases, a trader can take the other side of a trade in a similar contract on another exchange (e.g., taking a long position in the CBOT gold contract to unwind the short position taken in the NYMEX gold contract). However, such a position is not equivalent to complete offset of the initial position in at least two respects: (1) The trader is required to post margin on both trades; and (2) the trader still faces the (presumably small) likelihood of clearinghouse default on either exchange.
24. See, e.g., Damgard (2002), Frucher (2006), Harris (2006).
25. One can think of the central clearinghouse as a more efficient means of collusion between imperfectly competitive, integrated exchange/clearinghouses. An analogy between exchange/clearinghouses and other distributors might be instructive. If two imperfectly

competitive manufacturers agreed to sell their products through a common retailer rather than through separate captive retailers, they could collude more effectively.

26. Analogously, DeMarzo, Fishman, and Haggerty (2005) show that nonprofit exchanges reduce their enforcement efforts in order to create profits for their owner-members.
27. More generally, it is well established that a regulated monopolist that is controlled by a firm to which it provides an input may alter its practices in order to increase the unregulated firm's profits. See Armstrong and Sappington (2006) for a recent discussion of the theoretical and empirical literature on this topic.
28. This seems to be the model preferred by many FCMs.
29. For an overview of these issues, see Menell and Scotchmer (forthcoming).
30. For example, Penick (2005) reports that for the three main U.S. futures exchanges, the majority of contracts are no longer trading three years after their introduction.

REFERENCES

Armstrong, M., and D. Sappington. 2006. "Regulation, Competition, and Liberalization," *Journal of Economic Literature* 44, no. 2: 325–366.

Baer, H., V. G. France, and J. T. Moser. 2005. "Opportunity Cost and Prudentiality: An Analysis of Collateral Decisions in Bilateral and Multilateral Settings," in A. Chen, ed., *Research in Finance*, 201–228. New York, JAI Press.

Chicago Mercantile Exchange. 2006. *An Introduction to Futures and Options*. Chicago: CME Market Education Department.

Damgard, J. 2002. "Restructure Clearing," *Futures Industry Magazine*. Available at http://www.futuresindustry.org/fi-magazine-home.asp?a=791.

DeMarzo, P. M., M. J. Fishman, and K. M. Haggerty. 2005. "Self-Regulation and Government Oversight," *Review of Economic Studies* 72, no. 3: 687–706.

DeFontnouvelle, P., R. P. Fishe, and J. H. Harris. 2003. "The Behavior of Bid-Ask Spreads and Volume in Options Markets During the Competition for Listings in 1999," *Journal of Finance* 58: 2437–2463.

Edwards, F. R. 1983. "The Clearing Association in Futures Markets: Guarantor and Regulator," *Journal of Futures Markets* 3: 369–392.

Frucher, M. 2006. "Bearish on Chicago," *Wall Street Journal*, November 20.

Harris, L. 2006. "Breaking the Futures Monopoly," Forbes.com, November 6. Available at: www.forbes.com/opinions/2006/11/03/options-monopoly-cboe-cme-oped-cx_lh_1106options.html.

Hendershott, T., and C. M. Jones. 2005. "Island Goes Dark: Transparency, Fragmentation, Liquidity Externalities and Multimarket Regulation," *The Review of Financial Studies* 18: 743–793.

Hills, B., D. Rule, S. Parkinson, and C. Young. 1999. "Central Counterparty Clearing Houses and Financial Stability," *Bank of England Financial Stability Review* 6: 122–134.

Holder, M., M. Tomas III, and R. Webb. 1999. "Winners and Losers: Recent Competition among Futures Exchanges for Equivalent Financial Contract Markets," *Derivatives Quarterly* 6: 19–32.

Kane, E. J. 1980. "Market Incompleteness and Divergences between Forward and Futures Interest Rates," *Journal of Finance* 35: 221–234.

Kroszner, R. S. 1999. "Can the Financial Markets Privately Regulate Risk? The Development of Derivatives Clearinghouses and Recent Over-the-Counter Innovations," *Journal of Money, Credit, and Banking* 31: 596-619.

Menell, P., and S. Scotchmer. Forthcoming. "Intellectual Property," in A. M. Polinsky and S. Shavell, eds., *The Handbook of Law and Economics*, vol. 2. UC Berkeley Public Law Research, Paper No. 741724. Available at http://ssrn.com/abstract=741424.

Moser, J. T. 2000. "Origins of the Modern Exchange Clearing House: A History of Early Clearing and Settlement Methods at Futures Exchanges," in L. G. Telser, ed., *Classics on Futures: An Anthology with Commentaries*. London: Risk Publications.

Penick, M. 2005. "Tables of Contract Survival." Available from authors on request.

Pirrong, C. 2000. "Technological Change, For-Profit Exchanges, and Self-Regulation in Financial Markets." Presented at American Law and Economics Association Meetings, New York, 7 May, 2000.

Rudd, A., and M. Schroeder. 1982. "The Calculation of Minimum Margin," *Management Science* 28, no. 12: 1368–1379.

ABOUT THE AUTHORS

Jim Moser is deputy chief economist at the Commodity Futures Trading Commission. His research agenda focuses on three questions: What is the contribution of derivatives to the real economy? What arrangements do private markets make to resolve problems arising from use of derivatives? And, what is the proper role of government in regulating derivatives activity? His previous industry positions were at the Federal Reserve Bank of Chicago as a research officer and at the Chicago Mercantile Exchange, where he was a senior director. He has held academic positions at Louisiana Tech, Michigan State University, and the University of Kentucky. He received his BS in economics from Virginia Commonwealth University in 1978 and Ph.D. in finance from Ohio State in 1986.

David Reiffen earned a BS in economics at SUNY – Albany, and his master's and doctoral degrees in economics at UCLA. He is the author of more than 20 published articles in peer-reviewed journals in both economics and finance. These papers cover topics related to vertical restraints, retail pricing behavior, market microstructure, and pharmaceutical economics. Prior to joining the CFTC, Dr. Reiffen worked for the U.S. Federal Trade Commission, the Securities and Exchange Commission, and the Treasury Department. He has also served as an adjunct professor at the George Washington University.

CHAPTER 20

Counterparty Credit Risk

JAMES OVERDAHL
Chief Economist, U.S. Securities and Exchange Commission

In principle, the term *counterparty credit risk* refers to the possibility of nonperformance by a contract counterparty in any environment involving a contractual relationship. In practice, the term refers to the possibility of a payment default by a counterparty to an over-the-counter (OTC) derivatives contract, such as a swap. Within the OTC environment, end users, dealers, and interdealer brokers must each measure and manage counterparty credit risk exposure. In addition, regulators rely on measures of counterparty credit risk as part of their determination of regulatory capital requirements for regulated firms. Although primarily an OTC concern, counterparty credit risk also must be managed in the exchange-traded environment between the exchange-affiliated clearinghouse and clearing members and between clearing members and their customers. Counterparty credit risk is a concern in forward contracting relationships, such as between dealer banks in the interbank foreign exchange market or between a farmer and a grain elevator in the corn or soybean market. Counterparty credit risk also arises in securities financing transactions, such as repurchase agreements or securities lending.

This chapter provides a broad overview of the terms and concepts used for measuring and managing counterparty credit risk. First we describe the basic concepts used to measure both current counterparty credit risk (i.e., the current value of the exposure if a default occurred today) and potential counterparty credit risk (i.e., the value of the potential future exposure that may occur over the life of a transaction). These measures apply to individual positions, a portfolio of positions with a single counterparty, and a portfolio of positions across multiple counterparties. Then we describe the methods used by participants in derivatives markets to manage their counterparty credit risk exposure. These methods include risk-mitigating contractual provisions such as collateral agreements and close-out netting agreements. Recent initiatives to improve counterparty credit risk management by enhancing the infrastructure supporting OTC transactions, including central counterparty clearing, are discussed next.

The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The views expressed herein are those of the author and do not necessarily reflect the views of the Commission or of the author's colleagues upon the staff of the Commission.

MEASURING COUNTERPARTY CREDIT RISK EXPOSURE

Presettlement versus Settlement Risk

Participants in derivatives market use several terms to describe specific aspects of counterparty credit risk. At the broadest level, a distinction is made between presettlement risk and settlement risk. Presettlement risk is the credit risk arising from the possibility that default can occur prior to the final settlement of cash flows associated with a derivatives contract. Settlement risk is the risk of default when the contract matures or reaches a settlement date where the counterparties are required to exchange payments.[1] In the discussion that follows, only presettlement risk is considered.

Counterparty credit risk is best understood in the context of a swap transaction. Both parties to a swap contract face counterparty risk, but the risk of loss occurs only when a party to the contract is owed money. For example, consider a fixed-for-floating interest rate swap that began its life a year ago as a fairly priced (i.e., zero-value) swap and currently has four years of remaining life. Over the past year, short-term interest rates have gone up, meaning that the swap is now more valuable to the floating rate receiver (who is also the fixed rate payer) and less valuable to the floating rate payer (who is also the fixed rate receiver). Based on current market conditions, the mark-to-market value of the swap is $5 million to the floating rate receiver and, correspondingly, minus $5 million to the floating rate payer. If the swap is terminated today, the floating rate payer owes the floating rate receiver $5 million. In other words, because his swap position is in-the-money and he is owed money, the floating rate receiver is exposed to counterparty credit risk because there is a possibility the counterparty (the floating rate payer) may not be able to pay the $5 million he owes. The floating rate payer, whose swap position is out-of-the-money, currently faces no counterparty credit exposure because his counterparty owes him nothing.

As the example shows, the loss from a default today, assuming zero recovery, is the current mark-to-market value of the swap to the in-the-money counterparty (the counterparty to whom the value of the swap is positive). The example also demonstrates that either party to a swap contract could potentially hold the in-the-money position and therefore be currently exposed to counterparty credit risk.

Replacement Cost, Current Exposure, and Potential Exposure

The loss represented by the current mark-to-market value is equivalent to the cost of replacing the swap with a new swap under current market conditions. For this reason, the term *current replacement cost* or *current exposure* (CE) is used to describe counterparty credit exposure assuming default and contract termination occurs today with zero recovery. The current replacement cost is the larger of zero and the market value of a swap (for an individual swap transaction) or the larger of zero and the market value of a portfolio of swaps (for a portfolio of swaps). This exposure can be larger if the position is large and/or illiquid.

Current replacement cost alone does not accurately portray the potential credit risk over the life of the swap. A counterparty may default at some future date with swap values significantly different than current swap values. The potential loss is larger because the replacement cost can potentially become larger over the life of the swap. A counterparty acquires exposure not only due to the current value of the swap, but also due to potential changes in swap value. A potential change in exposure over the contract's life is called potential future exposure (PFE).

Simulation Techniques and the Exposure Profile

As a first step to measuring potential future exposure, end users and swap dealers estimate potential replacement cost for each of the swaps they hold with individual swap counterparties.

Measuring potential future exposure relies on simulation techniques. Simulation techniques rely on probability distributions to project the range of potential values of the swap over the swap's remaining life. Of interest are only the positive values the swap can potentially take on, since it is these values that are meaningful in the event of a default. For simulations far out into the future, there can be much uncertainty about potential swap value resulting in a wide range of estimates.

One common simulation technique is Monte Carlo simulation. Monte Carlo simulation is used to simulate a variety of different scenarios for potential value of a swap contract or a portfolio of swap contracts. The approach requires assumptions about the distribution of daily changes in the underlying market factors, such as interest rates, foreign exchange rates, or market indexes that affect the value of a swap or a portfolio of swaps. A common assumption in applying Monte Carlo simulation is that changes in the underlying market factors are normally distributed. Using historical observations of changes in the factors, the parameters of each factor distribution is estimated. Monte Carlo simulation then uses the estimated factor distribution to generate a simulated set of possible future daily changes in the factors. For each set of simulated factor changes, the swap or portfolio of swaps is revalued using valuation models, resulting in the swap being "marked to the future." The result is a set of simulated values for the swap or portfolio of swaps corresponding to the set of simulated changes in the underlying market factors. The upper bound of the set of simulated values represents the highest measure of counterparty credit exposure likely to occur at each date in the future.

The major determinants of potential credit exposure for a swap position are the time remaining in the life of the swap contract, and the volatility of the underlying market variables that determine the swap's value. The first determinant, called the amortization effect by Smithson, Smith, and Wilford (1995), captures the idea that credit exposure declines as the remaining life of the contract declines. This is because, over time, as more and more settlement payments are made, there are fewer remaining cash flows at risk. The second determinant, called the diffusion effect, captures the idea that the number of potential values for the variables determining the swap's value increase with the tenor of the swap.

When the series of credit exposure measures is plotted against the swap's tenor, a time-varying potential future exposure profile for the counterparty is produced. This profile is called the counterparty PFE profile. Exhibit 20.1 shows that, together, the amortization and diffusion effects lead to an exposure profile over the life of a swap that rises and then falls. The rising part of the profile shows that the diffusion

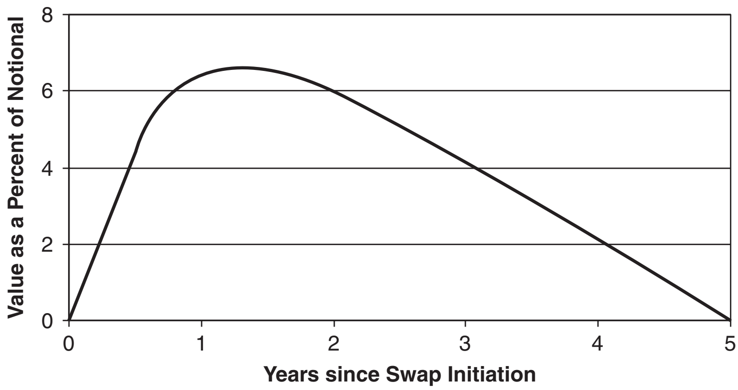


Exhibit 20.1 Potential Future Exposure Profile for a 5-Year Swap

effect dominates early in the life of the swap. The falling part of the profile shows that at some point in the swap's life the amortization effect comes to dominate.

It is common for swap participants to characterize counterparty credit risk with a single number. Projecting potential exposures over the life of the swap will produce many numbers. For example, a swap with five years' remaining life will have credit exposures measured for various potential default dates: for example, one year in the future, two years in the future, and so on. To reduce credit exposure to a single number, swap market participants frequently choose the maximum potential future exposure (MPFE) (i.e., the greatest amount that would potentially be lost over the life of the swap in the event of a counterparty default). The maximum exposure is a probability estimate of the "true" maximum exposure. Another frequently used measure is expected exposure (EE) that is simply the average future exposure calculated to a particular date, or maturity bucket that is a subset of the swap's remaining life. Exhibit 20.2 identifies the maximum potential future exposure and the expected exposure from the potential future exposure profile.

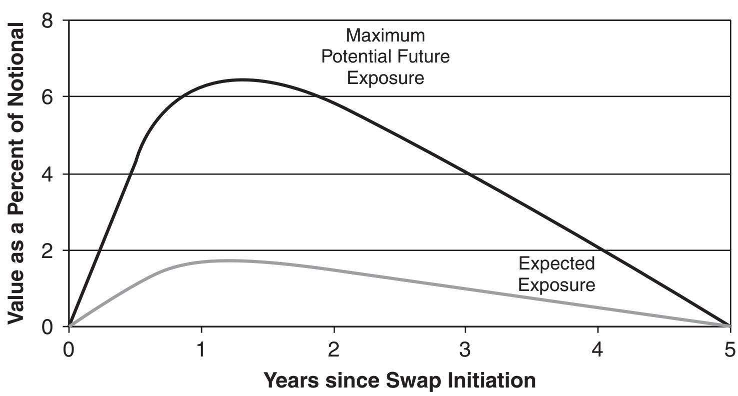


Exhibit 20.2 Maximum Potential Future Exposure and Expected Exposure

Maximum potential future exposure is a conservative measure that is used primarily to allocate credit (i.e., to determine how much to allocate for the transaction against the counterparty's credit line). Because mark-to-market values will be stochastic, there can be no true maximum exposure. The counterparty must therefore define the maximum exposure with some specified probability (e.g., 90 percent). Dealers use expected exposure to assist in pricing a transaction or to assist in evaluating the profitability of a transaction.

Wrong-Way and Right-Way Risk

Participants in the OTC derivatives markets also pay attention to the correlation between the cash flows of a swap and the creditworthiness of the counterparty. Market participants use two terms to describe this correlation. The first type of correlation to consider is wrong-way risk. This risk describes the potential deterioration in the ability of a counterparty to pay as the value of the swap moves out-of-the-money to the counterparty. For example, an investor may seek credit protection on a collateralized default obligation (CDO) by entering into a credit default swap (CDS) with a financial guarantor, such as a monoline bond insurance company. If the value of the swap increases to the investor, she is not only faced with current counterparty credit exposure to the guarantor but may find that the creditworthiness of the guarantor has deteriorated as a result of the market move (the wrong-way risk). This risk arises from the possibility that the financial guarantor may have sold protection to several counterparties and may not be able to meet all obligations if all the credit default swaps simultaneously move in-the-money to the protection buyers (and out-of-the-money to the financial guarantor). With wrong-way risk, counterparty credit risk is positively correlated with the amount of the counterparty's potential obligation under the terms of the swap.

In some situations, it is possible that the ability of a counterparty to pay is positively correlated with the amount of a counterparty's potential exposure. This type of exposure is referred to as right-way risk. For example, suppose a bank is the fixed payer in a fixed-for-floating oil swap contract with an oil production company. The oil production company, as the floating rate payer, will pay the spot price of oil on the contract's payment dates. If the price of oil rises, the oil production company owes more to the bank (an increase in the bank's counterparty credit exposure), but the company's creditworthiness increases as the price of oil increases the company's profitability. With right-way risk, counterparty credit risk is negatively correlated with the amount of the counterparty's potential obligation under the terms of the swap. For a good description of wrong-way and right-way counterparty credit risk, see Pengelly (2008). Brigo and Chourdakis (2008) consider the effect of wrong-way and right-way risk in the valuation of credit default swaps.

MANAGING COUNTERPARTY CREDIT RISK

Evaluating the Creditworthiness of Counterparties

Participants in the OTC derivatives market manage their counterparty exposure in the first instance by carefully screening and selecting counterparties. Dealers who are not highly rated are effectively precluded from the market, as no end user participant would use them. Dealers also engage in credit assessments of their end user counterparties. If an end user participant has an internal credit rating that is believed to be below investment grade, then that participant will be required to provide credit enhancement in the form of a collateral agreement.

The use of collateral agreements has grown over time. Surveys conducted by the International Swaps and Derivatives Association (ISDA) show that at the end of 2007, in excess of $2.1 trillion in collateral had been posted to support OTC derivatives exposures, compared to $0.2 trillion in 2000. During this time, the number of collateral agreements grew from 12,000 to 149,000. About 63 percent of OTC derivatives trades and exposures were covered by a collateral agreement in 2007 compared to 30 percent in 2003 (ISDA 2008). The collateral posted has shifted over time. In 1998, government securities were the predominant form of collateral, whereas in 2008 cash is most frequently posted (about 78 percent, according to the ISDA Margin Survey 2008). Gibson (2006) estimates that collateral and margin agreements can potentially reduce counterparty credit exposure by over 80 percent.

A collateral agreement will specify the initial collateral as well as variation margin requirements. The variation margin requirement will specify the amount of collateral that will change hands between counterparties on any given day. The contract will define a number of features including the threshold level, below which no collateral will be held. The contract will specify a minimum transfer amount, below which no margin transfer is made. The remargining period—that is, the interval at which margin is monitored and called for—is also specified in the collateral agreement.

Although collateral agreements can reduce counterparty credit risk, they are not magic bullets. Collateral cannot eliminate all counterparty credit exposure because collateral agreements sometimes include uncollateralized thresholds, minimum transfer amounts, or delays in mark-to-market valuations and margin calls that lead to temporary uncovered exposures. In addition, the use of collateral can lead to legal risk and liquidity risk. These risks can constrain the risk mitigation potential from using collateral. Legal risk can arise for a variety of operational reasons and from disagreements regarding the population of trades covered by the collateral agreement (perhaps caused by novations to different legal entities) or regarding the valuation of complex OTC products. Liquidity risk can arise because collateralization can be a source of funding liquidity risk because the counterparties have to provide collateral at relatively short notice.

Using Counterparty Credit Risk Measures in the Trade Authorization Process

After a counterparty's creditworthiness has been assessed, either with or without a collateral agreement, the next step is for OTC market participants to estimate potential future exposure to the counterparty of a proposed transaction. This is done using the techniques described earlier to simulate potential future exposure in various market scenarios with various future dates. Being able to estimate potential future exposure is essential for being able to monitor, manage, and allocate counterparty credit line usage relative to the counterparty credit limits. Dealers and end users are particularly interested in monitoring the incremental effect of a new transaction on credit line usage resulting from an additional transaction with a counterparty. If a new transaction results in a total credit exposure to a counterparty that exceeds a predetermined level, then a trader will fail to receive permission to execute the transaction. In addition to use in the trade authorization process, the PFE often is used to determine economic and regulatory capital.

Other Tools to Manage Counterparty Credit Risk

Other tools are available to manage counterparty credit risk. One tool is to use credit default swaps or contingent credit default swaps to buy protection, from a third party, on the original counterparty. Liquidity puts, credit triggers, and other early-termination provisions also are used to reduce credit exposures by shortening the effective maturities of trades. Liquidity puts give the parties the right to settle and terminate trades on prespecified future dates. Credit triggers specify that trades must be settled if the credit rating of a party falls below prespecified levels.

Netting

A contentious issue in the evaluation of credit risk concerns what is known as netting. The master agreements of most swap contracts state that if a counterparty defaults on one swap, it must default on all swaps. Therefore, swaps with a negative replacement cost will offset swaps with a positive replacement cost when the defaulting counterparty is the same in the two cases. Under such a netting arrangement, counterparty credit exposure is defined as the net positive replacement cost after netting.

To be effective, simulation methods must account for netting effects and bankruptcy rules. There is no reason that all swaps between dealers and end users will be subject to the same legal rules. Some swaps may be covered by a legally binding netting agreement, and others may not. The simulation method must apply the appropriate legal rules to each contract. In addition, the method must account for what a bankruptcy court would do in the event of a default. For example, in one legal scenario, a bankruptcy court may determine the replacement cost of each swap in the portfolio and simultaneously close out all positions. In an alternate legal scenario, a bankruptcy court may allow each swap to run until its settlement, maturity, or expiration date and then close out only those swaps that have positive replacement cost. Selectively closing out only those swaps with positive value is called cherry picking. The possibility of cherry picking is a scenario that must be accounted for in measuring potential counterparty credit exposure. Enforceable close-out netting provisions are now regarded as essential components to bankruptcy law in most financial market jurisdictions around the world.

Actual Default Experience in the OTC Market

Swap-related credit losses historically have been low, at least in the United States. According to statistics compiled by the Office of the Comptroller of the Currency (OCC), defaults peaked in the third quarter of 1998 (the time of Asian/Russian meltdown) at about $400 million, or about 12 percent of dealers' combined credit

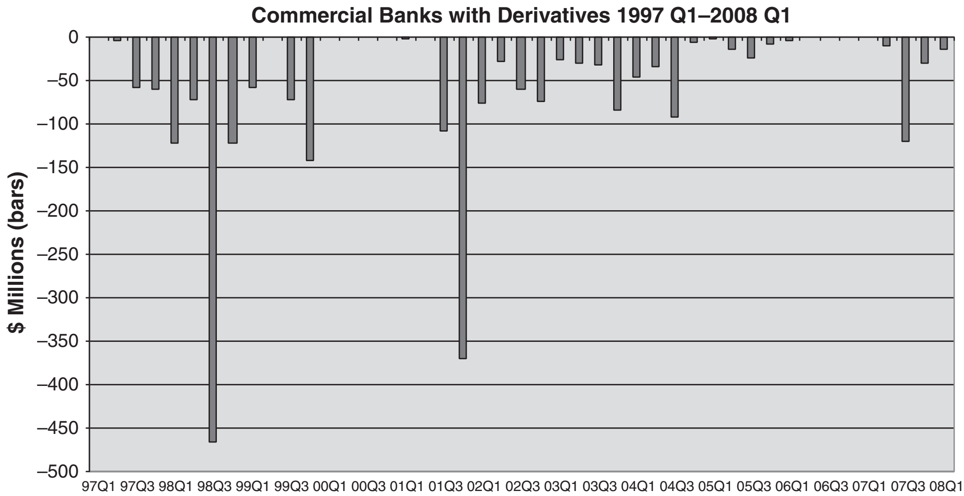


Exhibit 20.3 Quarterly Charge-offs from Derivatives: Commercial Banks with Derivatives 1997 Q1–2008 Q1

exposure. Typically, defaults are much lower. Exhibit 20.3 shows charge-offs for derivatives at U.S. commercial banks as reported in quarterly bank call reports.

INFRASTRUCTURE IMPROVEMENTS AIMED AT MITIGATING COUNTERPARTY CREDIT RISK

Infrastructure and the Effectiveness of Counterparty Credit Risk Management

As the market for OTC derivatives has grown, several regulators and market participants have expressed concerns that the infrastructure for clearing and settling OTC transactions has not kept pace with this growth.[2] As Bernanke (2008) and others have noted, weaknesses in infrastructure for clearing and settlement has left the financial system potentially vulnerable to the bankruptcy of a major counterparty, such as what arguably happened with the collapse of Lehman Brothers in September 2008. These weaknesses include the posttrade processing of OTC transactions. For example, large documentation backlogs of unconfirmed trades grew to high levels by 2005, especially for credit default swaps and equity swaps. This documentation backlog occurred not only for newly initiated transactions but also for novated contracts (i.e., contracts where one counterparty takes the place of another). Master agreements require prior written consent of the original counterparty to process a novation. However, the CRMPG II report (2005) demonstrated that dealers frequently accepted novations without prior consent. A backlog in documentation for novated contracts can lead to confusion as to the identity of the contract's counterparties. The backlog in documentation and confirmations arose because as the market grew the posttrade processing infrastructure remained decentralized and paper based.

At first blush, the quality of the infrastructure for clearing and settlement may seem to be an uninteresting part of the plumbing of financial systems, with no obvious connection to counterparty credit risk. However, backlogs in unconfirmed trades and errors in trade matching can undermine the effectiveness of counterparty credit risk management by allowing errors in trade records that can lead counterparties to underestimate their exposure or fail to collect margin when due. The backlog can also lead to market risk when, for example, a trader with an unconfirmed position thinks he has a hedge in place with a counterparty but is mistaken. This mistake, when discovered, means that the trader will have to replace the hedge at the prevailing market rate, which can be costly.

In 2008, the President's Working Group on Financial Markets (PWG) and the Financial Stability Forum (FSF) asked regulators to improve arrangements for clearing and settling credit default swaps and other OTC derivatives. Specifically, they asked for regulators to work at reducing trade processing backlogs and errors by insisting that the industry adopt high standards for trade data submission and the resolution of trade-matching errors. They also asked regulators to ensure that the industry adopt a cash settlement protocol for credit default swaps to address concerns that a large-scale default, or a large number of defaults occurring at the same time, could lead to a disorderly market if the industry relied on physical settlement. The PWG and the FSF also recommended that regulators ask the industry to develop an automated operational infrastructure for OTC derivatives that covers all major product types and market participants.

Central Counterparty Clearing

One industry initiative to improve the infrastructure of OTC markets has been the development of central counterparty (CCP) clearing. In the energy sector, the New York Mercantile Exchange, through its ClearPort facility, and the IntercontinentalExchange, through its ICE Clear facility, began to offer OTC clearing in the wake of credit disruptions caused by the Enron bankruptcy in 2001. In London, SwapClear, a service offered by LCH.Clearnet Limited, clears almost 50 percent of global single-currency interest rate swaps between dealers (Parkinson, 2008). The Chicago Mercantile Exchange offers a clearing facility, called CME Clearing 360, for designated interest rate swaps and foreign exchange swaps. In 2008, ICE Trust, which is part of the Intercontinental Exchange, Bclear, which is part of the NYSE-Euronext Group, the CME Group Inc., in a joint venture with Citadel Investment Group LLC, and LCH.Clearnet announced plans for implementing CCP clearing for credit default swaps.

A CCP typically novates bilateral trades so that it assumes any counterparty risks. Novation allows the CCP to enter into separate contractual arrangements with both of the initial counterparties—becoming buyer to one and seller to the other. A CCP has the potential to reduce counterparty risks by facilitating the multilateral offset and netting of obligations arising under contracts that are cleared through the system. A CCP also allows for anonymous trading, which means that market participants can execute transactions without worrying about the creditworthiness of the ultimate counterparty. The CCP becomes the counterparty to all market participants. Anonymous trading and settlement can avoid the signaling of distress that can occur absent a CCP, as bilateral counterparties novate trades away from one counterparty to be replaced by another. A CCP can also reduce counterparty credit risk by its ability to implement sound risk management practices. Sound risk management at a CCP is essential because a CCP concentrates risk. As part of its risk management, a CCP may subject novated contracts to initial and variation margin requirements or establish a clearing fund. The CCP also may implement a loss-sharing arrangement among its participants to respond to a participant insolvency or default. Thus, a CCP can serve a valuable function in reducing systemic risk by preventing the failure of a single market participant from having a disproportionate effect on the overall market. International standards on risk management practices for CCPs were agreed to in 2004 by the Committee on Payment and Settlement Systems of the central banks of the Group of 10 countries and the Technical Committee of the International Organization of Securities Commissions.

While providing a number of potential benefits, a CCP for credit derivatives or any OTC derivatives contracts is subject to substantial challenges. This is because the markets for OTC derivatives are generally less liquid than markets for exchange-traded instruments. As a result, the traditional procedures for a CCP to handle a default may not be as effective for these products. The traditional procedures for handling a default, which are used by CCPs for most exchange-traded derivatives, call for the CCP to terminate all of its contracts with the defaulting participant and promptly enter the market and replace the contracts. This will hedge against further losses on the open positions created by termination of the defaulter's contracts. But if the markets for the contracts cleared by the CCP are illiquid or under stress, entering the market may induce adverse price movements, especially if the defaulting participant's positions are large. Consequently, the application of traditional default procedures to illiquid OTC contracts may entail significant risk to the CCP.

Accordingly, a CCP should not be regarded as a silver bullet for concerns about the management of counterparty credit exposures related to OTC derivatives. Even with a CCP, preventing a systemic risk buildup would require dealers and other market participants to manage their remaining bilateral exposures effectively and the dealers' management of their bilateral exposures would require ongoing supervisory oversight. Nonetheless, developing a CCP for OTC derivatives would be an important step in accomplishing this goal. Pirrong (2008–9) analyzes the likely impact on systemic risk resulting from a move to central counterparty clearing for credit default swaps.

CONCLUSION

This chapter provides a broad overview of the measurement and management of counterparty credit risk in OTC markets. As the market for OTC derivatives has matured, so has the analysis of counterparty credit risk. The references provide a good foundation for gaining an in-depth understanding the state-of-the-art thinking about the measurement and management of counterparty credit risk. However, the infrastructure for documenting, processing, clearing, and settling these transactions has not kept up with the growth of the market. Recent government and industry initiatives are aimed at improving this infrastructure in order to facilitate more effective management of counterparty risks.

ENDNOTES

1. Settlement risk is sometimes referred to as Herstatt risk after a default episode in June 1974 involving the German bank Bankhaus Herstatt. In this episode, funds owed to Bankhaus Herstatt had been transferred to the bank through the interbank settlement process used to settle foreign exchange contracts. However, before transferring the funds it owed, Bankhaus Herstatt was closed due to insolvency. The resulting chain reaction of defaults roiled the inter-bank foreign exchange market.
2. See the President's Working Group on Financial Markets (2008), the Financial Stability Forum (2008), the Bank for International Settlements Committee on Payment and Settlement Systems (2007), and the Counterparty Risk Management Policy Group II (2005).

REFERENCES

Ammann, M. 2001. *Credit Risk Valuation: Methods, Models and Applications*, 2nd ed. Berlin: Springer.

Arvanitis, A., and J. Gregory. 2001. *Credit: The Complete Guide to Pricing, Hedging and Risk Management*. London: Risk Books.

Bank for International Settlements, Committee on Payment and Settlement Systems. 2007. "New Developments in Clearing and Settlement for OTC Derivatives" (March).

Bernanke, B. S. 2008. "Financial Regulation and Financial Stability." Speech at the Federal Deposit Insurance Corporation's Forum on Mortgage Lending for Low and Moderate Income Households, July 8. Arlington, VA.

Brigo, D., and K. Chourdakis. 2008. "Counterparty Risk for Credit Default Swaps," Working paper, Fitch Solutions, May.

Canabarro, E., and D. Duffie. 2003. Measuring and Marking Counterparty Risk, in L. Tilman, ed. *Asset/Liability Management of Financial Institutions*. London: UK Euromoney Books.

Canabarro, E., E. Picoult, and T. Wilde. 2003. "Analysing Counterparty Risk," *Risk* 16, no. 9 (September): 117–122.

Counterparty Risk Management Policy Group, "Towards Greater Financial Stability: A Private Sector Perspective. The Report of the CRMPGII," July 27, 2005.

Financial Stability Forum. 2008. "Enhancing Market and Institutional Reliance." Report to the G-7 Ministers and Central Bank Governors, April 11.

Gibson, M. S. 2006. "Measuring Counterparty Credit Exposure to a Margined Counterparty," in Michael Pykhtin, ed., *Counterparty Credit Risk Modelling: Risk Management, Pricing and Regulation*. London: Risk Books.

International Swaps and Derivatives Association Margin Survey. 2008. The following URL is available: http://www.isda.org/c_and_a/pdf/ISDA-Margin-Survey-2008.pdf

Lucas, D. H. 1995. "The Effectiveness of Downgrade Provisions in Reducing Counterparty Credit Risk," *Journal of Fixed Income* 5, no. 1 (June): 32–41.

Parkinson, P. M. 2008. Testimony before the Subcommittee on Securities, Insurance, and Investment, Committee on Banking, Housing, and Urban Affairs. "Reducing Risks and Improving Oversight in the OTC Credit Derivatives Market," United States Senate, July 9.

Pengelly, M. 2008. "Going the Wrong Way," *Risk* (March): 28–30.

Picoult, E. 1996. "Measuring Pre-Settlement Credit Risk on a Portfolio Basis," in *Risk Measurement and Systemic Risk: Proceedings of a Joint Central Bank Research Conference*. Washington, DC: Board of Governors of the Federal Reserve System.

Picoult, E., and D. Lamb. 2004. "Economic Capital for Counterparty Credit Risk," in Ashish Dev, ed., *Economic Capital: A Practitioner Guide*. London: Risk Books.

Pirrong, C. 2008–9. "The Clearinghouse Cure," *Regulation* (Winter): 44–51.

President's Working Group on Financial Markets. 2008. "Policy Statement to Improve Future State of Financial Markets," March 13.

Smithson, C., C. Smith Jr., and D. Sykes Wilford. 1995. *Managing Financial Risk: A Guide to Derivative Products, Financial Engineering, and Value Maximization*. Burr Ridge, IL: Irwin.

Sorensen, E. H., and T. F. Bollier. 1994. "Pricing Swap Default Risk," *Financial Analysts Journal* 50, no. 3 (May/June): 23–33.

ABOUT THE AUTHOR

James Overdahl is chief economist of the U.S. Securities and Exchange Commission. Prior to joining the SEC, he served as chief economist of the Commodity Futures Trading Commission. He received his B.A. degree from St. Olaf College and his Ph.D. in economics from Iowa State University. He has written extensively on market microstructure, risk management, and investments, and is co-author, with Robert Kolb, of three textbooks, *Financial Derivatives*, *Understanding Futures Markets*, and *Futures*, *Options and Swaps*.

CHAPTER 21

The Regulation of U.S. Commodity Futures and Options

WALTER L. LUKKEN
Commissioner and Former Acting Chairman, U.S. Commodity Futures Trading Commission

The Commodity Exchange Act (CEA), as amended,[1] is the statute authorizing the Commodity Futures Trading Commission's (CFTC) regulation of the commodity futures and options markets and their participants. The foundations of this act date back to 1922, when Congress brought agriculture futures into a modern regulatory structure.[2] The CEA generally requires that the trading of all futures contracts occur on a registered futures exchange[3] under the exclusive jurisdiction of the CFTC.[4] This regulatory structure originally was designed for traditional "open outcry" exchanges to ensure that futures trading occurred in specific physical locations under the watchful eye of a federal regulator. This regimented oversight also took aim at off-exchange "bucket shops," in which individuals would fraudulently solicit funds from customers for futures trading and then pocket, or "bucket," the money. Requiring that futures trading occur on a registered exchange among registered brokers greatly minimized this illegal activity.

The statutory basis for federal regulation of futures trading recognizes that these transactions are "entered into regularly in interstate and international commerce" and provide "a means for managing and assuming price risks, discovering prices, or disseminating pricing information through trading in liquid, fair and financially secure trading facilities."[5] Based on this public mission, the CFTC is charged with responsibility for detecting and preventing price manipulation, ensuring the financial integrity of all transactions, protecting market participants from fraud and other abusive trading practices, and promoting responsible innovation and fair competition.[6] These statutory purposes reflect the historic balance the CFTC must seek in preventing wrongful activity in the markets while encouraging fair competition and innovation in the industry.

This philosophy is also reflected in the most significant rewriting of the commodities laws in the last 25 years, the Commodity Futures Modernization Act of 2000 (CFMA).[7] This legislation amended the CEA to create a principles-based, tiered regulatory structure and design a more predictable and risk-based regulatory apparatus for derivative instruments, depending on the nature of the products being traded and the sophistication of the market participants. This regulatory structure was further amended by the recent passage of the CFTC Reauthorization Act as part of the Food, Conservation and Energy Act of 2008 (Farm Bill)[8] and is likely to be significantly amended as a result of the financial crisis of 2008.

TIERED REGULATORY DESIGN

The CFMA provided a new regulatory structure for the futures exchanges and their participants. The law amended the CEA to lay out different tiers of regulation for exchanges, depending on the types of products being traded and the level of sophistication of the participants trading them. Futures on commodities of finite supply that are more susceptible to manipulation and are offered to the retail public, such as agricultural futures contracts, are more heavily regulated under the designated contract market"(DCM) category.[9] Those instruments that are less susceptible to manipulation and are offered only to sophisticated investors and institutions would benefit from a lighter regulatory touch under the derivatives transaction execution facility (DTEF) registration.[10] The law also allowed exchanges that might otherwise qualify for an exclusion from the act to opt to be an exempt board of trade (EBOT), the lowest tier of regulation by the CFTC.[11] The intent behind the tiered regulatory structure was to allow exchanges as well as market participants to innovate more rapidly to meet competitive challenges while enabling the CFTC to tailor its regulatory focus to those areas requiring greater government scrutiny.[12]

The CFMA transitioned the regulatory structure of the CFTC from a prescriptive rules-based approach to one using principles and guidance, establishing the CFTC as the first U.S. federal financial regulator to adopt this structure. Instead of specifying the means for achieving a specific statutory mandate, the CFMA set forth core principles that are meant to allow participants in these markets to use different methodologies in achieving statutory requirements.

In fleshing out these core principles, the law allowed the CFTC to issue "best practice" regulations for complying with the core principles and allowed the CFTC to approve such acceptable business practices as are submitted to the CFTC for consideration.[13] The CFTC ultimately retains the authority to approve such practices and also may develop guidance of its own accord.

The CFMA also provided exchanges with authority to approve new products and rules through a self-certification process.[14] In self-certifying a new rule or product, the exchanges must provide the CFTC with a written declaration that the new contract or rule complies with the act and the CFTC's regulations.[15] The CFMA also reserved for exchanges the ability to request CFTC approval of a new rule or contract prior to becoming effective.[16] The CFMA further required prior agency approval for rule amendments that materially affect futures on enumerated agricultural commodities.[17] These changes were seen as necessary to allow exchanges the ability to react quickly to the competitive challenges anticipated by the CFMA.[18]

The CFMA required for the first time that clearinghouses receive their own separate designation as derivatives clearing organizations (DCOs) before providing clearing services.[19] Like exchanges, a DCO must abide by a separate set of core principles tailored to the specific risks associated with these entities.[20] The CFMA also enabled registered DCOs to clear excluded over-the-counter (OTC) derivatives in an effort to reduce counterparty and systemic risks for these transactions.[21] Following the demise of Enron, the New York Mercantile Exchange (NYMEX) sought and received approval from the CFTC in 2002 to clear OTC energy products for the first time. Today, a significant number of OTC energy derivatives are cleared through regulated clearinghouses, which has reduced systemic risk and allowed regulators a greater window into this marketplace. Clearing for OTC products now extends beyond energy products to financial products such as credit default swaps, forward rate agreements, and foreign currency swaps.

STATUTORY EXCLUSIONS FOR CERTAIN OTC DERIVATIVES

As amended by the CFMA in 2000, the act excludes certain OTC derivatives products from the CFTC's jurisdiction based on the nature of the underlying commodity and the types of participants in the markets. The economic functions of OTC derivatives and futures contracts are similar, as both seek to transfer undesired economic risk from one party to another willing to accept it. Unlike standardized futures contracts, however, many OTC derivatives are not traded on registered futures exchanges, have individually tailored terms, and are privately designed and brokered for sophisticated counterparties by sophisticated financial institutions.

Prior to the CFMA, the CEA requirement that futures contracts be traded on-exchange created a degree of uncertainty about the legality of these off-exchange OTC instruments, causing systemic risk concerns among policy makers. To address this situation, and based largely on recommendations of the President's Working Group on Financial Markets (PWG),[22] Congress enacted in the CFMA several statutory exclusions from CFTC oversight for certain OTC derivatives. In determining whether the CFTC should have regulatory authority, the CFMA looked to whether the products are being traded by retail customers, whether the products are susceptible to price manipulation, and whether the participants are not otherwise regulated.

The CFMA incorporated this approach in several provisions. In the highly liquid financial marketplace, the CFMA excluded from the CFTC's jurisdiction financial products that are traded among large and/or regulated entities, defined as eligible contract participants (ECPs), as long as the trades are not transacted on an exchange-like trading facility.[23] The definition of ECPs includes financial institutions, insurance companies, investment companies, commodity pools, corporations, pension plans, governmental entities, broker-dealers, futures commission merchants, floor brokers and traders, and individuals with total assets in excess of $10 million.[24]

The CFMA also excluded from CFTC oversight financial products traded on an electronic trading facility if the transactions are entered into on a principal-to-principal basis by ECPs.[25] In addition, the CFMA excluded from CFTC jurisdiction the electronic trading facilities upon which these products are traded.[26] Electronic exchanges were thought to be less susceptible to wrongdoing due to the real-time audit trail created by these entities. By prohibiting brokered trades as a condition to this exclusion, it was believed that participants trading for their own accounts would be more responsible for their actions and subject to greater market discipline.[27]

The CFMA further excluded hybrid instruments that are predominantly securities, providing a four-part test for determining whether transactions qualify for the exclusion.[28] In addition, the CFMA excluded individually negotiated nonagricultural swaps that are entered into by ECPs and not executed on an exchange-like trading facility.[29] This provision, which arguably overlaps with other exclusions and exemptions contained in the CFMA, was meant to ensure that conventional bilateral OTC swaps were outside the jurisdiction of the CFTC.

As a final legal certainty measure, the CFMA provided that OTC derivatives transactions among ECPs should not be void simply due to a failure of one of the conditions of the exclusion.[30] This language aimed to ensure that institutions would not walk away from their OTC obligations due to a technical contractual defect.

SECURITY FUTURES PRODUCTS

The CFMA also lifted the statutory ban on security futures products (SFPs), which had been in place since 1983, and provided a joint regulatory framework between the Securities and Exchange Commission (SEC) and CFTC for allowing the trading of these instruments. The CEA provides the CFTC exclusive jurisdiction over futures on broad-based security indices.[31] For futures on single securities or narrow-based indices, known collectively as SFPs, the CFMA provided the CFTC and SEC with joint jurisdiction over these instruments as both futures and securities.[32] The CFMA and subsequent regulations defined "narrow-based security index" to mean an index that has nine or fewer securities or in which the component securities are weighted in a manner that could allow the manipulation of a single or small group of securities within the index.[33] To avoid duplicative regulation, the CFMA established a system of "notice registration" under which trading facilities and intermediaries that are already registered with either the CFTC or the SEC may register on an expedited basis with the other agency for the limited purpose of trading SFPs. Such notice-registered entities are fully regulated by the primary regulator and subject to limited regulatory requirements imposed by the secondary notice regulator. The trading of SFPs has met with limited success in the United States.

The CFMA grandfathered existing equity futures contracts that were trading at the time of its enactment, including several foreign security index futures contracts. This statutory provision was reflected in a June 2002 order issued by the agencies, allowing existing futures on foreign security indices to continue trading in the United States. For nongrandfathered foreign products, the CFMA set a deadline for the SEC and CFTC to develop joint rules for futures on foreign broad-based security index products by December 21, 2001. It also provided that the agencies develop joint rules on foreign narrow-based products. As of the date of publication, the agencies have yet to issue joint rules on these foreign products.

In March 2008, the CFTC signed a mutual cooperation agreement with the SEC to establish a closer working relationship between the agencies, establish a permanent regulatory liaison between the agencies, provide for enhanced information sharing, and establish several key principles guiding the agencies' consideration of novel derivative products that may reflect elements of both securities and commodity futures or options.[34] This agreement led to the expedited, coordinated approval of the trading and clearing of several novel derivative products (futures and option contracts based on shares of certain exchange-traded funds), an outcome expected to enhance legal and regulatory certainty for users of these novel products.

RETAIL FOREIGN CURRENCY FRAUD

Prior to the CFMA's enactment, the CEA excluded certain foreign currency transactions from CFTC jurisdiction as long as they did not involve contracts for future delivery "conducted on a board of trade." This ambiguous statutory language, known as the Treasury Amendment because it had been inserted at the behest of the Department of Treasury during the CEA's enactment in 1974, had been subject to conflicting judicial interpretations that made it difficult for the CFTC to bring fraud actions against off-exchange foreign currency futures scams aimed at retail customers.[35]

To clarify this jurisdictional uncertainty, the CFMA adopted language generally excluding from the CFTC's oversight foreign currency transactions not transacted on a registered futures exchange. However, where off-exchange retail foreign currency futures or options transactions are offered, the CFMA clarified that the CFTC does have jurisdiction unless the offering firm is an "otherwise regulated" entity (such as a bank, a broker-dealer, a financial or investment bank holding company, or an insurance company). The CFMA further clarified that the CFTC's antifraud authorities apply to off-exchange foreign currency transactions that futures commission merchants (FCMs) and their affiliates enter into with retail customers. These changes temporarily plugged the loophole that had allowed off-exchange retail foreign currency bucket shops and boiler rooms to flourish.

Since 2000, the CFTC has used this authority to aggressively pursue retail foreign currency fraud through enforcement actions and cooperation with local, state, and federal criminal authorities. The CFTC has shut down hundreds of firms, frozen assets, and assisted in criminal prosecutions that have led to significant jail time for violators.

In 2004, however, the Seventh Circuit Court of Appeals curtailed the CFTC's ability to combat retail off-exchange foreign currency fraud in the *Zelener* decision.[36] The court held that the contracts at issue in that case were not futures contracts, rejecting the CFTC's multifactor test for such determinations, and found that the transactions were a type of rolling spot contract not subject to the CFTC's jurisdiction. Likewise in 2008, the Sixth Circuit Court of Appeals in the *Erskine* case followed the *Zelener* reasoning in limiting the CFTC's ability to combat retail foreign currency fraud.[37] While both decisions concluded that fraud clearly had been committed, the courts dismissed the cases as not properly within the CFTC's jurisdictional reach. As will be discussed, the CFTC Reauthorization Act of 2008 amended the CFTC's authority over retail foreign exchange (forex) transactions to address many of these issues.

EXEMPT COMMERCIAL MARKETS

Exempt commercial markets (ECMs) have been the subject of considerable public attention recently given historic high prices in the energy sector. In 2000, the CFMA provided legal and regulatory certainty for certain exempt OTC commodity transactions and markets. The CFMA defined "exempt commodity" as a commodity that is "not an excluded commodity or an agricultural commodity."[38] In practice, this definition primarily encompasses energy and metal commodities. Unlike the other statutory *exclusions* where the CFTC retains no authority or jurisdiction over transactions, an *exemption* retains for the CFTC certain residual authorities while serving to clarify the areas of the law that no longer pertain to a given transaction. In the CFMA, this exemptive language stated that nothing in the act applies to a transaction in an exempt commodity, which is entered into by ECPs and not traded on a trading facility,[39] except that the CFTC retains certain antifraud and antimanipulation authorities over these exempt transactions.[40]

In addition, the CFMA provided an exemption from the CFTC's jurisdiction, found in §2(h)(3) of the CEA, for transactions in exempt commodities traded on an electronic trading facility as long as they are entered into on a principal-to-principal basis among eligible commercial entities (ECEs).[41] An eligible commercial entity is generally defined as an ECP that is either a large dealer or a commercial participant in the commodity business.[42] As part of the §2(h)(3) exemption, the transactions are subject to certain antifraud and antimanipulation authorities of the CFTC.[43] The conditioned exemption also requires an ECM relying on the exemption to notify the CFTC of its intention to operate and the names of its owners; to describe the types of commodity categories being traded; to identify its clearing facility, if any; to certify that the facility will comply with the terms of the exemptions; and to certify that the owners of the trading facility are not otherwise statutorily disqualified under the act.[44] The ECM has six other requirements; it must:

1. Either provide the CFTC with real-time access to its trading system and protocols, or provide the CFTC with reports on specific trading positions.
2. Maintain books and records for five years.
3. Agree to provide the CFTC with specific information on a special call basis.
4. Agree to submit to the agency's subpoena authority.
5. Agree to comply with all applicable laws and require the same of its participants.
6. Not represent that the facility is registered or in any way recognized by the CFTC.[45]

As will be discussed, the CFTC Reauthorization Act of 2008 further amended the CFTC's authority over ECMs to provide the agency with greater authority when ECM contracts serve a significant price discovery function.

CFTC REAUTHORIZATION ACT OF 2008

On May 22, 2008, the CFTC Reauthorization Act of 2008 was enacted into law, making critical improvements to the CEA and the CFTC's authority. The new legislation reauthorized the CFTC through FY 2013, closed the so-called Enron

Loophole by allowing enhanced CFTC oversight of ECMs that trade contracts linked to regulated U.S. futures contracts, increased CFTC penalty authority for manipulation and false reporting, clarified the CFTC's antifraud authority for off-exchange principal-to-principal energy trades, and clarified the CFTC's retail foreign currency antifraud authority.

The collapse of Enron and the implosion of the Amaranth hedge fund highlighted certain regulatory gaps that had developed in the energy trading sector—commonly known as the Enron Loophole—and the potential need for additional regulatory authorities over OTC energy transactions. Based on a CFTC study and report on ECMs in the fall of 2007,[46] Congress amended the CFTC's authorities for ECMs as part of the Farm Bill. Congress provided that, upon a determination by the CFTC that an ECM futures contract serves a significant price discovery function, the CFTC has four additional authorities:

1. To require large trader position reporting for that contract
2. To require an ECM to adopt position limits or accountability levels for that contract
3. To require an ECM to exercise self-regulatory responsibility over that contract in preventing manipulation
4. To exercise emergency authority (along with the ECM) over that contract

On March 23, 2009, the CFTC finalized its rules and regulations for these new ECM authorities.

In addition, based largely on the recommendations of the PWG, the new authority requires those who participate in the solicitation of retail forex transactions to register with the CFTC. It also closes the loophole that allowed firms to notice register with the SEC as securities broker-dealers—under the CFMA's SFP provisions—and then serve as counterparties to retail off-exchange forex transactions. Last, the legislation bolstered the CFTC's enforcement authority over retail off-exchange forex transactions like those in dispute in the *Zelener* and *Erskine* cases and required entities that serve as dealers in these markets to maintain $20 million in capital. These amendments to the CEA are designed to strengthen the CFTC's enforcement powers against fraudulent retail foreign currency activity, and CFTC rulemakings for these provisions are expected in mid 2009.

FUTURE LEGISLATIVE REFORMS

In the wake of the financial crisis of 2008, the new Congress and administration has proposed broad financial reform of the regulatory structure. There have been growing calls for wholesale regulatory changes and consolidation, including initiatives to scrap the entire financial regulatory system for an objectives-based approach or to simply merge the CFTC and SEC.[47] These long-term structural issues are likely to be a focus of the new Congress and administration through 2009, but reform efforts will face difficult substantive and jurisdictional obstacles that have prevented such restructuring in the past.

Volatility in the energy and agricultural markets will continue to be a focal point of policy makers as they work to ensure the proper functioning of the price discovery and risk management futures markets. With the rise of crude oil and agricultural commodity futures prices to record highs in 2008, there have been many legislative proposals to limit the role of speculators and passive investors in the futures markets, ranging from outright bans to position limits to higher margin requirements. In addition, Congress is also likely to turn to legislation to bring more transparency and oversight to OTC derivatives. Most of these efforts seek to move a large segment of the standardized OTC derivatives markets onto regulated exchanges and clearinghouse with the remaining non-standardized contracts subject to prudential regulation and heightened capital charges. These legislative initiatives will continue to be discussed and debated as part of the larger financial regulatory reform efforts in 2009 and 2010.

ENDNOTES

1. Commodity Exchange Act, 7 U.S.C. §§ 1–27 (1994 & Supp. 2003).
2. The original act, known as the Grain Futures Act, was enacted September 21, 1922. In 1936, the amendments to the act changed the name to the Commodity Exchange Act and broadened the agricultural commodities covered by the law.
3. CEA § 4(a).
4. Id. § 2(a)(1)(A). See also Chicago Mercantile Exchange v. SEC, 883 F.2d 537 (7th Cir. 1989).
5. CEA § 3.
6. Id.
7. Commodity Futures Modernization Act of 2000, Pub. L. No. 106-554, 114 Stat. 2763 (December 21, 2000).
8. The Food, Conservation and Energy Act, Pub. L. No. 110-246, 122 Stat. 1651 (June 2008).
9. CEA § 5.
10. Id. § 5a.
11. Id. § 5d.
12. At the end of 2007, there were 12 DCMs and 8 EBOTs. To date, there have been no DTEFs.
13. CEA § 5c(a).
14. Id. § 5c(c).
15. Id. § 5c(c)(1).
16. Id. § 5c(c)(2)(A).
17. Id. § 5c(c)(2)(B).
18. Others have recognized the potential competitive benefits of the self-certification process for the securities industry. See the Department of the Treasury Blueprint for a Modernized Financial Regulatory Structure, March 2008 at 116; available at: www.treas.gov/press/releases/reports/Blueprint.pdf
19. CEA § 5b. At the end of 2007, there were 11 DCOs.
20. Id. § 5b(c)(2).
21. Id. § 5b(b).
22. Report of the President's Working Group on Financial Markets: Over-The-Counter Derivatives Market and the Commodity Exchange Act, November 1999 (PWG Report).
23. CEA § 2(d)(1).
24. Id. § 1a(12).
25. Id. § 2(d)(2).
26. Id. § 2(e).
27. See PWG Report at 18.
28. CEA § 2(f).
29. Id. § 2(g).

30. Id. § 22(4).
31. Id. § 2(a)(1)(C)(ii).
32. Id. § 2(a)(1)(D)(i).
33. Id. § 1a(25).
34. See www.cftc.gov/stellent/groups/public/@newsroom/documents/file/cftc-sec-mou030608.pdf.
35. See, e.g., CFTC v. Frankwell Bullion Ltd., 99 F.3d 299 (9th Cir. 1996).
36. CFTC v. Zelener, 373 F.3d 861 (7th Cir. 2004).
37. CFTC v. Erskine, 512 F.3d 309 (6th Cir. 2008).
38. CEA § 1a(14).
39. Id. § 2(h)(1).
40. Id. § 2(h)(2)(B).
41. At the end of 2007, 19 ECMs had notified the CFTC of their intention to rely on the §2(h)(3) exemption.
42. Id. § 1a(11).
43. Id. § 2(h)(4).
44. Id. § 2(h)(5).
45. Id.
46. See www.cftc.gov/stellent/groups/public/@newsroom/documents/file/pr5403-07_ecmreport.pdf.
47. For example, see the Department of the Treasury Blueprint for a Modernized Financial Regulatory Structure, March 2008; available at: www.treas.gov/press/releases/reports/Blueprint.pdf, and the Group of Thirty Report on the Structure of Financial Supervision (October 2008).

ABOUT THE AUTHOR

Walter Lukken was first appointed commissioner of the U.S. Commodity Futures Trading Commission (CFTC) in 2002 and is now serving his second term due to expire in 2010. President Bush nominated him in September 2007 to serve as chairman of the CFTC. He served as acting chairman of the Commission from June 27, 2007, until January 20, 2009. Commissioner Lukken has testified numerous times before Congress and has represented the agency as part of the President's Working Group on Financial Markets. He also has represented the Commission before international organizations and forums, including the International Organization of Securities Commissions and the Committee of European Securities Regulators. Prior to joining the CFTC, Commissioner Lukken served for five years as counsel on the professional staff of the U.S. Senate Agriculture Committee, specializing in futures and derivatives markets. He received his BS degree with honors from the Kelley School of Business at Indiana University and his Juris Doctor degree from Lewis and Clark Law School in Portland, Oregon.

CHAPTER 22

Accounting for Financial Derivatives

IRA G. KAWALLER
President, Kawaller & Co., LLC.

Most commercial enterprises contemplating the use of derivatives need to understand both the economics of these instruments (i.e., how they work, how they are priced, etc.) and the accounting treatment that will be applied. This chapter addresses that second concern.

Rules for accounting for derivatives and hedging transactions are found in the Financial Accounting Standard No. 133 (FAS 133), a standard that was originally issued in 1998 but has been amended numerous times. Besides these amendments, additional guidance has also been promulgated in the form of Derivative Implementation Group (DIG) Issues, which are posted to the Financial Accounting Standards Board (FASB) web site (www.fasb.org). Together, this guidance often is cited as the most complex of all accounting standards; consequently, this chapter should be read only as a high-level overview. Readers who have "special circumstances" are advised that the information presented in this chapter may be incomplete or inappropriate for their particular situations.

ALTERNATIVE ACCOUNTING CATEGORIES

The cardinal accounting rule is that derivatives must be recognized as assets or liabilities and recorded on the balance sheet at their fair market value. The question of where *changes* in value are posted (i.e., to earnings or elsewhere on financial statements) is the nub of the issue; and the answer depends on (1) how the derivative is being used and (2) whether prerequisite conditions have been satisfied to allow for "special hedge accounting."

In the case when derivatives are used for trading purposes, or, alternatively, if no special hedge designation has been documented, the derivatives' gains or losses are recorded in current earnings, as they arise. In other words, derivatives are carried on the balance sheet at fair market value, with changes posted to earnings. In other cases—that is, when designated as hedges, assuming the qualifying criteria are satisfied—different accounting follows depending on the nature of the exposure being hedged. Three different types of hedge treatments are specified under FAS 133: Cash flow hedges, fair value hedges, and hedges of net investments in foreign operations.

Importantly, hedge accounting is not automatic. Rather it is an elective, assuming strict criteria are satisfied. The motivation for seeking this treatment is that it assures that the earnings impacts of the derivative are recognized in earnings concurrently with those associated with the exposure being hedged, thereby allowing for the income statement to more faithfully reflect the economic objective of using derivatives in the first place. Otherwise (i.e., without hedge accounting), the income statement would likely reflect a considerably higher degree of earnings volatility, as the earnings of these two sides of the hedging relationship would impact earnings in separate accounting periods.

To qualify for hedge accounting of any type, the hedge relationship must be documented completely and correctly as of the date that the hedge accounting is to be applied. Put another way, if this documentation is not in place at the time the derivative position is initially transacted, hedge accounting would not be appropriate. Once properly documented, however, hedge accounting could be started from that point.

Depending on the nature of the hedging relationship, certain aspects of hedge documentation may seem to be quite onerous, particularly if companies use derivatives that are not *precisely* tailored to address the risk that they are intended to hedge. In such cases, companies need to articulate prospective and retrospective effectiveness tests in their documentation. These tests must demonstrate that, while not perfect, the hedges are expected to be "highly effective" prospectively; and, in fact, this expectation must be validated *ex post*. Critically, the notion of being highly effective is not necessarily that the instrument performs as planned but rather that the gains or losses generated by the derivative closely offset the losses or gains associated with the risk being hedged. Perhaps surprisingly, much of the difficulty (and controversy) relating to hedge accounting is that merely performing as advertised does not necessarily mean that the desired offset can necessarily be expected to occur.

When these offsets are expected to be other than perfect, the documentation must offer effectiveness test designed to show that the offsets will be (prospectively) and are (retrospectively) within acceptable bounds. The problems are twofold: (1) the test design must reflect the facts and circumstances of the hedge relationship in question, and (2) sometimes it is harder to pass these tests than you might expect. In any case, hedge accounting can be applied *only* if and when the documentation is properly articulated and the effectiveness tests are satisfied.

Cash Flow Hedges

A hedge of an upcoming, forecasted event is a cash flow hedge. In this case, the exposure being hedged must involve the risk of an uncertain (i.e., variable) cash flow. Derivative results must be evaluated, with a determination made as to how much of the result is effective and how much is ineffective. The ineffective component of the hedge results must be posted to current income, while the effective portion is initially recorded in other comprehensive income (OCI) and later reclassified to income in the same time frame in which the forecasted cash flow affects earnings.

For purposes of determining the amount that is appropriate to be posted to OCI, the assessment must be made on a cumulative basis. Contributions to earnings are required only if the derivative results exceed the cash flow effects of the hedged items under the current guidance, but this requirement may be subject to change. Before a change becomes generally accepted accounting practice (GAAP), an exposure draft of the proposed change has to be circulated, and the reporting public will have an opportunity to react to the proposed changes. As of this writing, no change is expected imminently. Still, the FASB has set the process in motion.

Importantly, qualifying for hedge accounting requires an ongoing assessment. Criteria must be satisfied both at the inception of the hedge and throughout the term of the hedging relationship. Thus, if after initially qualifying for cash flow accounting the criteria for hedge accounting stop being satisfied, hedge accounting would no longer appropriate. With the discontinuation of hedge accounting, any accumulated OCI would remain there, however, unless (except in extenuating circumstances) it is probable that the forecasted transaction will not occur by the end of the originally specified time period or within an additional two-month period of time thereafter (Paragraph 33 of FAS 133).

Reporting entities have complete discretion that allows for dedesignating cash flow hedge relationships at will and later redesignating them, assuming all hedge criteria are again (or still) satisfied (Paragraph 32c).

Examples of Exposures that Qualify for Cash Flow Hedge Accounting

- Interest rate exposures that relate to a variable or floating interest rates.
- Planned purchases or sales of assets.
- Planned issuances of debt or deposits.
- Planned purchases or sales of foreign currencies.
- Currency risk associated with prospective cash flows that are not denominated in the functional currency.

In all of these bulleted points, the common theme is that the company is facing a prospective transaction whereby the amount paid or received is uncertain and thereby subject to risk.

Eligible Risks (Paragraphs 29g and 29h)

- Currency risk associated with:
  a. A forecasted transaction in a currency other than the functional currency.
  b. An unrecognized firm commitment.
  c. A recognized foreign currency–denominated debt instrument.
- The entire price risk associated with purchases or sales of nonfinancial goods. That is, unless the purchase or sale specifically relates to buying or selling individual components, the full price of the good in question must be viewed as the hedged item.
- For interest-bearing instruments, hedgeable exposures include cash flow effects to:
  a. Changes in the full price of the instrument in question.
  b. Changes the benchmark rate of interest (i.e., the risk-free rate of interest or the rate associated with London Interbank Offered Rate [LIBOR]–based swaps).

c. Changes associated with the hedged item's credit spread relative to the interest rate benchmark.
d. Changes in cash flows associated with default or the obligors' creditworthiness.
e. Changes in currency exchange rates (Paragraph 29h).

Prerequisite Requirements to Qualify for Cash Flow Accounting Treatment

- Hedges must be documented at the inception of the hedge, with the objective and strategy stated, along with an explicit description of the methodology used to assess hedge effectiveness (Paragraph 28a).
- Dates (or periods) for the expected forecasted events and the nature of the exposure involved (including quantitative measures of the size of the exposure) must be explicitly documented (Paragraph 28a).
- The hedge must be expected to be highly effective, both at the inception of the hedge and on an ongoing basis. Effectiveness measures must relate the gains or losses of the derivative to changes in the cash flows associated with the hedged item (Paragraph 28b).
- The forecasted transaction must be probable (Paragraph 29b).
- The forecasted transaction must be made with a different counterparty than the reporting entity (Paragraph 29c).

Disallowed Situations (i.e., when Cash Flow Accounting May Not Be Applied)

- In general, written options may not serve as hedging instruments. An exception to this prohibition (i.e., when a written option may qualify for cash flow accounting treatment) is when the hedged item is a long option (Paragraph 28c).
- In general, basis swaps do not qualify for cash flow accounting treatment unless both of the variables of the basis swap are linked to two distinct variables associated with two distinct cash flow exposures (Paragraph 28d).
- Cross-currency interest rate swaps do not qualify for cash flow hedge accounting treatment if the combined position results in exposure to a variable rate of interest in the functional currency. This hedge *would* qualify, however, as a fair value hedge.
- With held-to-maturity fixed income securities under Statement 115, interest rate risk may not be designated as the risk exposure in a cash flow relationship (Paragraph 29e).
- The forecasted transaction may not involve a business combination subject to Opinion 16 and does not involve:
  a. A parent's interest in consolidated subsidiaries;
  b. A minority interest in a consolidated subsidiary;
  c. An equity-method investment; or
  d. an entity's own equity instruments (Paragraph 29f).
- Prepayment risk may not be designated as the hedged item (Paragraph 29h).
- The interest rate risk to be hedged in a cash flow hedge may not be identified as a benchmark interest rate, if a different variable interest rate is the specified exposure—for example, if the exposure is the risk of a higher prime rate, LIBOR may not be designated as the risk being hedged (Paragraph 29h).

Internal Derivatives Contracts

- Except in the case when currency derivatives are used in cash flow hedges, derivatives between members of a consolidated group (i.e., internal derivatives) cannot qualify as hedging instruments in the consolidated statement, unless offsetting contracts have been arranged with unrelated third parties (Paragraph 36).
- For an internal currency derivative to qualify as a hedging instrument in a consolidated statement, it must be used as a cash flow hedge only for a foreign currency forecasted borrowing, a purchase or sale, or an unrecognized firm commitment, but the following conditions apply:
  - The nonhedging counterpart to the internal derivative must offset its net currency exposure with a third party within three days of the internal contract's hedge designation date (Paragraph 40).
  - The third-party derivative must mature within 31 days of the internal derivative's maturity date (Paragraph 40).

Fair Value Hedges

When hedging exposures associated with the price of an asset, a liability, or a firm commitment, the total gain or loss on the derivative is recorded in earnings. In addition, the underlying exposure due to the risk being hedged must also be marked to market to the extent of the change due to the risk being hedged; and these results flow through current income as well. This treatment is called a fair value hedge. Hedgers may elect to hedge all or a specific identified portion of any potential hedged item.

As with cash flow hedges, fair value hedges also requires specific criteria to be satisfied both at the inception of the hedge and on an ongoing basis. If, after initially qualifying for fair value accounting, the criteria for hedge accounting stop being satisfied, hedge accounting is no longer appropriate. With the discontinuation of hedge accounting, gains or losses of the derivative (assuming it has not been liquidated) will continue to be recorded in earnings, but no further basis adjustments to the original hedged item would be made (Paragraph 26).

Again, as with cash flow hedges, reporting entities have complete discretion to de-designate fair value hedge relationships at will and later redesignate them, assuming all hedge criteria remain (Paragraph 24).

Examples of Exposures that Qualify for Fair Value Hedge Accounting

- Interest exposures associated with value changes of fixed rate debt.
- Price exposures for fixed rate assets.
- Price exposures for firm commitments associated with prospective purchases or sales.
- Price exposures associated with the market value of inventory items.
- Price exposures on available-for-sale securities.

As mentioned, the hedged items in fair value hedges are either recognized assets or liabilities or firm commitments, subject to the requirement that the instrument under consideration is subject to price risk. With this sensitivity in mind, it is useful to distinguish between fixed rate and variable rate debt securities. While the price of fixed rate debt varies inversely with interest rates, variable rate debt maintains its value at par (at least as of any settlement dates). Thus, with respect to interest-bearing securities, fair value hedging would apply *only* to fixed rate instruments.

Eligible Risks (Paragraph 21f and 36)

- The risk of the change in the overall fair value.
- The risk of changes in fair value due to changes in the benchmark interest rates (i.e., the risk-free rate of interest or the rate associated with LIBOR-based swaps), foreign exchange rates, creditworthiness, or the spread over the benchmark interest rate relevant to the hedged item's credit risk.
- Currency risk associated with:
  a. An unrecognized firm commitment;
  b. A recognized foreign currency–denominated debt instrument;
  c. An available-for-sale security.

Prerequisite Requirements to Qualify for Fair Value Accounting Treatment

- Hedges must be documented at the inception of the hedge, with the objective and strategy stated, along with an explicit description of the methodology used to assess hedge effectiveness (Paragraph 28a).
- The hedge must be expected to be highly effective, both at the inception of the hedge and on an ongoing basis. Effectiveness measures must relate the gains or losses of the derivative to those changes in the fair value of the hedged item that are due to the risk being hedged (Paragraph 20b).
- If the hedged item is a portfolio of similar assets or liabilities, each component must share the risk exposure, and each item is expected to respond to the risk factor in comparable proportions (Paragraph 21a).
- Portions of a portfolio may be hedged if they are:
  a. A percentage of the portfolio;
  b. One or more selected cash flows;
  c. An embedded option (provided it is not accounted for as a stand-alone option);
  d. The residual value in a lessor's net investment in a direct financing or sale-type lease (Paragraph 21a2 and 21f).
- A change in the fair value of the hedged item must present an exposure to the earnings of the reporting entity (Paragraph 21b).
- Fair value hedge accounting is permitted when cross-currency interest rate swaps result in the entity being exposed to a variable rate of interest in the functional currency

Disallowed Situations (i.e., when Fair Value Accounting May Not Be Applied)

- In general, written options may not serve as hedging instruments. An exception to this prohibition (i.e., when a written option may qualify for cash flow accounting treatment) is when the hedged item is a long option. FAS 133 also defines any combinations that include a written option and involves the net receipt of premium—either at the inception or over the life of the hedge—as a written option position (Paragraph 20c).

- Assets or liabilities that are remeasured with changes in value attributable to the hedged risk reported in earnings—for example, nonfinancial assets or liabilities that are denominated in a currency other than the functional currency—do not qualify for hedge accounting. The prohibition does not apply to foreign currency–denominated debt instruments that require remeasurement of the carrying value at spot exchange rates (Paragraphs 21c, 29d, and 36).
- Investments accounted for by the equity method do not qualify for hedge accounting (Paragraph 21c).
- Equity investments in consolidated subsidiaries are not eligible for hedge accounting (Paragraph 21c).
- Firm commitments to enter into business combinations or to acquire or dispose of a subsidiary, a minority interest, or an equity method investee are not eligible for hedge accounting (Paragraph 21c).
- A reporting entity's own equity is are not eligible for hedge accounting (Paragraph 21c).
- For held-to-maturity debt securities the risk of a change in fair value due to interest rate changes is not eligible for hedge accounting. Fair value hedge accounting may be applied to a prepayment option that is embedded in a held-to-maturity security, however, if the entire fair value of the option is designated as the exposure (Paragraph 21d).
- Prepayment risk may not be designated as the risk being hedged for a financial asset (Paragraph 21f).
- Except for currency derivatives, derivatives between members of a consolidated group cannot be considered to be hedging instruments in the consolidated statement, unless offsetting contracts have been arranged with unrelated third parties on a one-time basis (Paragraph 36).

Hedges of Net Investments in Foreign Operations

Special hedge accounting is appropriate for hedges of the currency exposure associated with net investments in foreign operations, which give rise to translation gains or losses under Statement of Financial Accounting Standards No. 52. These are gains and losses that feed into the company's capital via an account called the currency translation account (CTA), without being reflected in the firm's income statement. Both derivatives and nonderivatives (e.g., liabilities denominated in the same currency as that of the net investment) may be designated as hedges of these exposures. Effective results of such hedges are recognized in CTA, coincident with the recognition of the net investment gains or losses. Ineffective portions of hedge results are recognized in earnings (Paragraph 42).

Again, specific criteria must be satisfied both at the inception of the hedge and on an ongoing basis. If and when the criteria for hedge accounting are no longer satisfied, hedge accounting must be discontinued; and with the discontinuation of hedge accounting, gains or losses of the derivative—if that derivative is still retained—will be recorded in earnings. Reporting entities have complete discretion to hedge relationships at will and later redesignate them, assuming all hedge criteria remain satisfied.

Prerequisite Requirements to Qualify for Hedge Accounting Treatment

- Hedges must be documented at the inception of the hedge, with the objective and strategy stated, along with an explicit description of the methodology used to assess hedge effectiveness. This documentation must include the identification of the hedged item and the hedging instrument and the nature of the risk being hedged (Paragraph 20a).
- The hedge must be expected to be highly effective, both at the inception of the hedge and on an ongoing basis. Effectiveness measures must relate the gains or losses of the derivative to those changes in the fair value of the hedged item that are due to the risk being hedged (Paragraph 20b).

CONCLUSION

The challenge in accounting for derivatives is that the required procedures depend not on the nature of the instrument but rather on the manner in which the derivatives are used. Beyond that, the procedures are complicated by the fact that special hedge accounting (which results in derivatives' gains or losses being reflected concurrently with the income effects of the associated hedged item) is not available simply on an election basis. Rather, reporting entities must specifically qualify for this treatment, and the assessment of whether they are qualified must be made on an ongoing basis. As a consequence, the accounting treatment may change over the life of the derivative. Thus, hedge accounting may apply during some portion of the derivative's holding period and not at other times.

Nobody said it was easy.

REFERENCES

FASB Statement No. 133 (incorporating FASB Statements Nos. 137, 138, and 149). 2004. *Accounting for Derivative Instruments and Hedging Activities*. Financial Accounting Standards Board, February 10.

Kawaller, I.G., and R. Steinberg. 2007. "Guide for Preparing Hedging Documentation," Accounting and Disclosure for Derivative Instruments, BNA Accounting Policy & Practice Portfolio 5112 (Worksheet 7).

ABOUT THE AUTHOR

Ira G. Kawaller is the managing partner of the Kawaller Fund and president of Kawaller and Co., the former being a commodity pool and the latter being a consulting company that assists companies in their use of derivatives. Prior positions were with the Chicago Mercantile Exchange, J. Aron & Company, AT&T, and the Board of Governors of the Federal Reserve System. Dr. Kawaller received a PhD in economics from Purdue University and has held adjunct professorships at Columbia University and Polytechnic University. Currently he is a board member of Hatteras Financial, a publicly traded real estate investment trust, and has served on a variety of professional boards and committees in the past, including the board of the International Association of Financial Engineers and the Financial Accounting Standard Board's Derivatives Implementation Group.

CHAPTER 23

Derivative Scandals and Disasters

JOHN E. MARTHINSEN
Professor of Economics and International Business, Babson College

INTRODUCTION

Derivative contracts have existed for more than 4,000 years (Swan, 2000). Invented to solve practical problems, these financial instruments have broadened and deepened our financial markets, made them more efficient, and, as a result, contributed materially to worldwide economic growth, development, and wealth creation. Despite their net contribution to the depth and breadth of our financial markets, there is a dark side in the abuse and misuse of derivative instruments that is almost as old as the markets themselves. Looking back on the past 35 years, when the derivatives markets, as we know them today, really took flight, two things are clear: Derivative disasters have grown explosively in size, and the speed with which these losses have materialized has accelerated.

Derivative scandals and disasters are so idiosyncratic that it is a challenge to find common threads that run through all of them. In this chapter, we take a look at five of the most sensational and controversial mishaps during the period from 1993 to 2008 (see Exhibit 23.1) to identify important similarities and the lessons we have learned.[1]

ANATOMY OF DERIVATIVE-RELATED FAILURES

Every derivative fiasco has its own particular death spiral, but four interrelated factors are common to all of them:

1. Exceptionally large wagers based on faulty strategies or the wild speculative activities of rogue traders.
2. Significant exogenous shocks that are difficult to predict and probabilistically rare.
3. Dysfunctional risk management systems.
4. The lack of access to reliable sources of liquidity when it is needed most.

Highly leveraged, speculative positions are at the root of virtually all derivative scandals and disasters. Access to deep pools of national and international liquidity

Exhibit 23.1 Five Major Derivative Scandals and Disasters: 1993 to 2008

Company	Abbreviation	Loss *(Billions)*	Year	Financial Instruments
Metallgesellschaft Refining and Marketing	MGRM	\$1.3	1993	U.S. oil and gas futures contracts
Barings Bank	Barings	\$1.2	1995	Japanese equity and government bond futures contracts
Long-Term Capital Management	LTCM	\$4.5	1998	Leveraged global spread trades
Amaranth Advisors	Amaranth	\$6.4	2006	U.S. natural gas futures contracts
Société Générale	SocGen	\$7.2	2008	European index futures

enables companies to finance assets at relatively cheap rates and then leverage their positions even further by investing in derivative instruments. Speculators are supposed to bring value to derivative markets through their ability to value and manage risks. Because efficient markets ensure that higher returns come only as a result of greater risks, the ability to target, measure, monitor, and control exposures is crucial to speculators' success. Failure in any of these areas can result in losses, and when the losses are grand enough, they become "financial disasters." See Exhibit 23.1 for examples of five major disasters between 1993 and 2008.

Derivative strategies are nothing more than theories that prescribe systematic ways to profit from *temporary* market anomalies. Normally, investors study historic trends and interrelationships to find prices that are perceived to be inconsistent with current or expected market fundamentals. When such inconsistencies occur, positions are taken, and profits are earned after fundamentals reassert themselves and prices revert to their fair market values. Profitability often is used to validate strategies, but success breeds confidence, and confidence nurtures a greater willingness increase exposures. As a result, successful traders can quickly build positions of considerable absolute and relative size.

Using the past to identify misaligned and unsustainable market values is most successful during extended periods of rising or falling prices. This technique is not as successful at estimating turning points or predicting the values of derivatives that have a paucity of historical data. But identifying mispriced risks is only half the job; predicting when prices will return to their expected levels is equally important. Timing often spells the difference between success and failure because holding unprofitable derivative positions requires capital, and capital has a cost. Leverage amplifies the losses on nonconverging positions, forcing decisions regarding whether to vacate, temporarily finance, or increase them. Without adequate financing, strategies that might become profitable with time must be abandoned, often at fire sale prices. It is for this reason that access to liquidity (i.e., reliable creditors, repo counterparties, and investors/shareholders) is essential, especially during periods of financial turmoil.

Performance-based compensation practices can add fuel to speculative fires by encouraging greater risk taking. Derivative traders and fund managers often are compensated with a mixture of relatively low, fixed, annual salaries and performance-based bonuses. These bonuses can provide entry to a potential compensation stairway to heaven because their payoff profiles are like long call options, with virtually unlimited upward potential but no downside risk. Bonuses are also like call options because they increase in value with volatility, but they are problematic because traders/fund managers can amplify this volatility by taking greater risks. If they win, traders/fund managers can earn thousands (even millions) of dollars in added compensation. If they lose, a degree of reputation might be lost, but their companies bear the financial burdens. Typically, unsuccessful traders are fired, change employers, or, in some cases (if you can believe it), set up their own hedge funds with eager investors at their doorsteps.

Unexpected exogenous shocks are responsible for igniting derivative disasters. In many cases, changes in market prices are so large that, by normal statistical standards, they should occur not even once in 1,000 years. With such diminutive odds, potential threats are not taken seriously, thereby blindsiding aggressive investment strategies when they occur. As losses mount, companies find themselves in desperate need of cash, but liquidity evaporates when news of financial troubles is communicated to the markets, causing creditors and counterparties to increase margin requirements, refuse to roll over credits, reduce credit ratings, seize and/or sell collateral, and aggressively mark positions to market. Investors/depositors react by withdrawing funds. Business lines are cut, and access to new credit is curtailed or eliminated. As a result, financing costs rise, and funding availability falls. During these tumultuous periods, panic often replaces measured judgment, causing the strained company to close positions in an untimely manner and inducing creditors and counterparties to recover as much as they can as quickly as they can.

Most of these companies start as enormous suppliers of liquidity to the derivative markets; so it is not at all surprising that a lack of liquidity becomes the ultimate coup de grâce, when they become significant demanders and try to unload positions in markets with no natural counterparties. This lack of access to sufficient liquidity often is exacerbated by plummeting market prices and uncertainty about the value of complex positions.

INVESTMENT STRATEGIES AND EXOGENOUS SHOCKS BEHIND OUR FIVE DERIVATIVE FIASCOS

What were the investment strategies of our five derivative-related scandals and disasters, and what exogenous shocks ignited these crises? Answers to these questions are important because they draw attention to the important gap between actual and perceived risks.

MGRM's Strategy

Metallgesellschaft AG (MG) was a German-based multinational company focusing on mining and smelting nonferrous metals. In 1991, the company diversified its activities by offering *forward* oil and gas contracts through Metallgesellschaft Refining and Marketing (MGRM), a newly established U.S. affiliate. MGRM's team of experienced traders offered customers long-term, fixed rate, forward energy contracts with maturities extending out as far as 10 years.[2] Customers were enthusiastic, and, by September 1993, MGRM had already established itself as the industry leader, supplying forward contracts for as many as 185 million barrels of oil to approximately 100 independent heating oil and gasoline retailers.

The enormous size of its short forward positions exposed MGRM to high levels of energy price risk. To ameliorate these exposures, the company used stack-and-roll hedges[3] to transform its energy *price risk* into energy *basis risk*.[4] Stack-and-roll hedges reduced the volatility of MGRM's long-term profitability, but, at the same time, their lopsided cash flows exposed the company to a considerable amount of liquidity risk.[5] MGRM willingly accepted both basis risk and short-term liquidity risk as part and parcel of its strategic business model. The company was confident that *temporary* cash outflows caused by unfavorable basis movements could be funded with the considerable financial resources of its German parent, MG, and *temporary* cash inflows could be invested wisely and efficiently.

MGRM was knocked off balance almost immediately after starting, during the early 1990s, when energy market conditions changed dramatically. International fears of oil shortages subsided when Saddam Hussein was defeated and Iraqi troops withdrew from Kuwait during early 1991. As a result, the spot price of oil fell relative to the forward price, causing the basis on energy contracts to move from backwardation (i.e., positive) to contango (negative). By 1993, the plummeting basis on energy contracts caused worrisome reductions in MGRM's reported profits, and declining prices caused the company's enormous stack of long futures positions to hemorrhage cash. MG's German-based supervisors panicked in late 1993, liquidated MGRM's hedges immediately, and thereby incurred sizable losses. These losses were compounded later when energy prices rose, causing MGRM's uncovered (short) forward positions to spill red ink.

LTCM's Strategy

LTCM's goal was to create a diversified, market-neutral portfolio that was based on small imperfections in the spread between the market prices of underliers (e.g., equities, bonds, currencies, and indices). When market fundamentals reasserted themselves and these imperfections vanished, LTCM expected to earn minuscule investment returns on each of its many positions, regardless of whether market prices rose or fell. The fund's operational goal was to raise the potential yearly fluctuations of its portfolio's net asset value to 20 percent. To accomplish this goal with its portfolio of low-risk, spread trades, LTCM borrowed considerable amounts, thereby increasing its asset-to-equity ratio (i.e., leverage).[6]

LTCM grew quickly to become a significant player in the global investment arena. The company was highly leveraged, with assets worth approximately $125 billion and equity equal to only about $5 billion. Added to this were thousands of open derivative positions, which had a total notional value equal to about $1.25 trillion.[7] In 1998, the fund held between 5 percent and 10 percent of the open interest on many international exchanges and transacted an even greater percent of the daily turnover on these exchanges.

To address potential liquidity risks during periods when market spreads were misaligned for prolonged periods, LTCM secured long-term funding through extended credit lines, unsecured loans, and term (reverse) repurchase agreements. It also developed a broad and comprehensive working capital model that provided incentives for traders to finance their positions using term agreements. To restrict investor outflows, LTCM wrote covenants into its investment contracts that locked up funds for extended periods and limited the number and size of withdrawals. Finally, to ensure that it could borrow at the lowest possible rates and have access to multiple lines of secure financing, LTCM kept its credit rating as high as possible.

The basis risk that LTCM faced was supposed to be mitigated by its diversified portfolio, but fund managers severely overestimated the level of diversification. As a result, LTCM's liquidity and equity needs proved to be far greater than anticipated. Company analysts had studied past asset correlations to find positions that would be relatively uncorrelated with each other. What they failed to realize soon enough was that LTCM's risk tolerances, funding needs, and liquidity requirements were similar to those of other (imitator) hedge funds. As a result, LTCM's positions were highly correlated with the hedge fund industry, and these correlations increased during turbulent financial periods (MacKenzie 2003, p 349–380).

From 1997 to 1998, the Asian Tiger Crisis (1997), Russian Ruble Crisis (1998), and turmoil in Indonesia, China, and the Middle East caused spreads between U.S. and global interest rates to widen as investors fled to safety. LTCM's diversified portfolio helped stem losses from these crises, but other hedge funds were not as fortunate. In an effort to finance their rapidly deteriorating positions, these funds dumped assets and unwound positions in a desperate flight to quality, flight to liquidity, and flight from arbitrage. As a result, price spreads for many of LTCM's major investments went helter skelter, causing massive losses.

Amaranth's Strategy

Amaranth began operations in 2000 as a multistrategy hedge fund. Its portfolio was not diversified, and it was not advertised to be. Rather, Amaranth altered its investment strategies to take advantage of the most promising markets and opportunities. Amaranth's investments were a combination of directional bets and spread trades (mainly calendar spreads). The leveraged and nondiversified structure of its portfolio increased Amaranth's exposure to market risk, which the fund accepted willingly as part of its business model. To control liquidity risk, the company relied on secure lines of funding, developed an internal cash management system, and also restricted investor withdrawals by using lockup periods, prewithdrawal notification rules, and gating provisions.

At first, the fund was heavily invested in convertible securities, but, in 2004, Amaranth began trading energy derivatives, intending to limit its exposures in this asset area to no more than 2 percent of its portfolio. Sagging profits in other sectors convinced Amaranth to enlarge these investments, and, by the end of 2006, energy investments had become 58 percent of Amaranth's portfolio. Amaranth's energy positions were massive by any standard. At times during 2006, the company held between 46 and 81 percent of the open interest in the most actively traded natural gas futures contracts traded on the New York Mercantile Exchange (NYMEX) for maturities extending as far out as 2010.[8]

The investment strategies that eventually led to Amaranth's failure began in January 2006. The company wagered that the U.S. natural gas industry would recover quickly from two deadly hurricanes (i.e., Katrina and Rita) in 2005 and create a glut of natural gas that would push down prices until the end of summer (2006). Thereafter, inclement U.S. weather conditions and production bottlenecks would eliminate the surfeit of natural gas by the winter 2006/2007 heating season, causing natural gas prices to rise until spring 2007. To profit from these expectations, Amaranth made major (calendar) spread bets that turned out to be sources of the fund's undoing. These bets were based on its expectation that winter 2007 natural gas prices would rise relative to both fall 2006 prices and spring 2007 prices.[9]

Amaranth's enormous (long) calendar spread positions would have earned handsome profits if the basis on natural gas contracts increased as the company predicted, but, during August and September 2006, they plunged as a result of relatively mild U.S. weather conditions and few serious hurricanes. Consequently, Amaranth's spread positions, which just a few months earlier had appeared to be highly profitable, became virtually impossible to liquidate at profitable prices and, eventually, had to be sold at considerable losses.

Barings' and Société Générale's Speculative Traders

The derivative-related tragedies at Barings in 1995 and SocGen in 2008 were different from our other financial disasters because they involved rogue traders (Nick Leeson at Barings and Jérôme Kerviel at SocGen), who were supposed to be arbitraging, executing orders on behalf of customers, and/or conducting legitimate trades for their banks. These traders were not authorized to take significant risks, and, therefore, they were not supposed to have independent strategies.[10] For this reason, inadequate risk management caused these financial disasters, not faulty company strategies.

Unlike the exposures at MGRM, LTCM, and Amaranth, which were related to basis risk, Barings and Amaranth faced virtually unfiltered directional price risk. Like wolves in sheep's clothing, Leeson and Kerviel made massive directional bets and then covered their tracks by entering fictitious trades in their banks' accounting systems that made it seem as if they were hedged.

During 1994 and early 1995, Leeson bet that the Japanese stock index (i.e., the Nikkei 225) and 10-year Japanese government bond index (JGB) would rise, but they fell, and their free fall was reinforced on January 17, 1995, by the Kobe earthquake, one of the worst natural disasters in Japan during the twentieth century. In 2008, Jérôme Kerviel bet that the European stock markets would rise, but they fell as the subprime crisis spread from the United States to financial and real markets around the world.

Lax supervision permitted these traders to accumulate positions that were far in excess of both their internal limits and the banks' shareholders' equity. Kerviel's gross positions grew to approximately €49 billion ($73 billion), which greatly exceeded SocGen's market capitalization of €34 billion ($52.6 billion) (Société Générale, 2008). Leeson's transactions were so large that, during February 1995, they looked like a single-handed effort to keep the Nikkei 225 Index from sinking below the 19,000 level. In February 1995, Barings' Singapore affiliate, Barings Futures (Singapore) (BFS), was counterparty to almost 9 percent of the trading volume on SIMEX, and it held 49 percent of the long open interest for SIMEX's March 1995 Nikkei 225 futures contract (Bank of England, 1995, Section 4.25).

LESSONS LEARNED FROM DERIVATIVE SCANDALS AND DISASTERS

The derivative-related financial fiascos from 1993 to 2008 taught us a number of important lessons. Six of them are:

1. Controlling risks is possible only if they can be effectively measured.
2. Risk management systems must react to cauterize losses immediately after an initial shock.
3. More creative ways are needed to supply liquidity to markets during turbulent times.
4. Strict oversight must be placed on individual traders and fund managers.
5. Compensation incentives and promotion criteria must be scrutinized to ensure they support company goals.
6. Risk management systems are only as strong as their weakest risk managers.

Controlling Risks Is Possible Only If They Can Be Measured Effectively

MGRM, LTCM, and Amaranth seriously underestimated the vulnerability of their portfolios to changing market conditions. As a result, they charged bargain prices for the risk protection they sold, which may have been part of the reason their businesses grew so swiftly. Their experiences point to a need for better pricing models that can more accurately value highly complex positions and provide more realistic appraisals during tranquil *and* rapidly changing periods. Therefore, considerable care must be taken to appraise and improve the quality of data from which valuation estimates are drawn. Pricing models need to be adaptive, with considerable flexibility to change assumptions and perform scenario analyses. Some companies have tried to validate their pricing models by selling small quantities of relatively illiquid assets to see how market prices compare to their estimates.

Stress tests and value at risk (VaR) analyses are frequently used tools for quantifying and monitoring financial exposures. Stress tests provide feedback on worst-case scenarios, and VaR analyses provide guidance for day-to-day operations. Sophisticated derivative users employ both measures in determining their optimal liquidity and equity levels. Inherent in VaR are some important assumptions, which, if violated, can cause mispriced risks. Specifically, VaR assumes that:

1. Asset prices change smoothly and continuously.
2. Asset returns are normally distributed and will fluctuate in the future as they have in the past.
3. Investors' decisions are independent of one another.
4. There will always be liquid markets into which assets can be sold, if needed.

These assumptions were violated in each of our five financial disasters. Market prices did not move continuously but instead changed in leaps and bounds. Returns were not normally distributed; rather, the distribution tails were fat, causing unexpected changes to occur more frequently than expected. Volatility did not remain at its historic average, and investors' actions were not independent. Finally, market liquidity dried up as perceived counterparty risk increased and accurate market valuations became more difficult to determine.

VaR indicated that LTCM should expect yearly losses no greater than $714 million (10.5 percent of its equity) for 99 of 100 *years*. But the fund lost $4.5 billion in fewer than *two months* during 1998. LTCM's average investment return (loss) turned out to be 6 standard deviations from its historical mean, which, statistically speaking, should have occurred not even once in a millennium. Similarly, Amaranth's VaR statistics indicated that the company's chances of losing more than 28 percent of its portfolio (i.e., about $2.9 billion) during any one month were about 1 in 100. Nevertheless, the company lost $6.4 billion in one month, with $4.6 billion of these losses occurring in just one week. This event was nine standard deviations from the historic mean return (see Till, 2006). If the returns were normally distributed, these price movements would have had a statistical chance of happening not even once in a million years.

One of the recurring themes with derivative disasters has been the apparent inability of market participants (e.g., creditors and counterparties) to correctly value a financially stressed company's portfolio *before* it fails but the miraculous speed with which these valuations can be done by potential buyers at the stroke of midnight on the day before bankruptcy would be triggered. This valuation puzzle is important because significant profits have been earned during very short periods from buyout deals. For example in September 2006, JP Morgan (JPM) and Citadel Investment Group LLC (Citadel) purchased Amaranth's portfolio and agreed to share the risks. The deal was finalized only after Amaranth agreed to make a concession payment of more than $2.5 billion to the buyers. Yet within two weeks of the agreement, JPM sold its positions to Citadel for a profit estimated to be at least $725 million, and, by that time, Citadel had hedged most of the risks associated with the positions it had already acquired from Amaranth.[11]

Risk Management Systems Must Cauterize Losses Immediately after the Initial Shocks

The derivative-related damages at all five of our companies could have been reduced considerably if their risk management systems had been able to stop or slow the losses immediately after the initial exogenous shocks. It was not first impact that caused these disasters but rather the companies' inability to limit losses after first blood was drawn. Among the major causes of losses at Barings, LTCM, Amaranth, and SocGen were doubling strategies used by traders in the weeks subsequent to the initial shocks.

Doubling occurs when a trader significantly increases the size of his or her bets (e.g., doubles them) in order to reverse a large loss or, perhaps, earn a small profit. This strategy is problematic because traders who employ it stand to lose considerable amounts in very short time periods. It is also a concern because, prior to the collapse, the individuals using doubling strategies often appear to be earning relatively healthy returns with low volatility (see Brown and Steenbeek, 2001, p. 84). Therefore, supervisors tend leave them alone for fear of interfering or meddling, but when the roof caves in, it collapses with such speed and force that recovery becomes virtually impossible.

Nick Leeson was so sure that he was right and the market was wrong that he increased his positions, from July 1994 to February 1995, with every drop in the Nikkei 225 index to take advantage of (what he thought were) cheap prices. The "Arb Boys" at LTCM (1998), Amaranth's Brian Hunter (2006), and SocGen's Jérôme Kerviel (2008) did the same as market prices and/or spreads moved decidedly against them. About three-quarters Barings' eventual \$1.3 billion loss occurred during the last month—almost half of it during the final week. LTCM incurred about half of its eventual \$4.5 billion loss during just one month from mid-August to mid-September 1998. More than 70 percent of Amaranth's \$6.4 billion loss occurred during just one week in September 2006, and SocGen suffered 100 percent of its \$7.2 billion loss during three weeks in January 2008, with about 70 percent occurring in the last three days.

Creative Ways Are Needed to Supply Liquidity during Turbulent Times

Access to sufficient liquidity is essential for companies with sound strategies that need time to harvest their returns. Two major sources of financing are from the outright sale of positions and from diversified funding lines. There is optimism in the potential for market-oriented solutions, such as the development of liquidity options, contingent capital agreements, reinsurance contracts, and incentives to fund assets with longer-term debt (see Scholes, 2000). Of course, these solutions come with costs that will surely reduce profitability, but, at the same time, they will also provide greater security against high-impact, difficult-to-predict, and probabilistically rare exogenous shocks.

The crucial significance of funding during financial crises has motivated hedge funds to insist on covenants in their contracts that curtail customer withdrawals, demand lockup periods, limit the number of customer withdrawals per year, require notice before withdrawing funds, and impose gating provisions that restrict withdrawals to no more than a certain percent of an investor's net asset value. Companies also use two-way collateral covenants and term agreements for posting collateral. Some funds require traders to invest a portion of their annual bonuses in the business, and these investments have vesting periods to prevent immediate withdrawals. Still other companies retain the right, under certain circumstances, to suspend redemptions. By reducing the risk of sudden runs, these provisions and covenants give companies more control of the liquidity needed to fund fluctuations in daily asset values.

Risk Management Systems Must Control Traders and Fund Managers

The financial disasters at Barings and SocGen made it obvious that that these banks had done little or nothing to control traders' intraday, overnight, and term positions, or to control the types of instruments traders could purchase and sell. Supervisors lost sight of the North Star of risk management, namely, incredibly large profits are more often the result of incredibly large risks than they are the Midas touch of extraordinarily talented employees.[12]

Derivative trades are made at the speed of light; so automated risk management systems are needed to track traders' daily activities. Internally developed risk management systems that are based on manual spreadsheets and black-box proprietary information need to be replaced with externally developed, automated systems that have impenetrable computer access codes (e.g., codes with multiple authentication factors, such as a password, cryptographic token, and fingerprint). They should deny write-access by front office employees to middle and back office applications, monitor who logs on and off the system, track open activity, and have algorithms that alert supervisors to idiosyncratic (fraudulent) trading behavior. Risk management systems have to ensure that controls are carried out on time, feedback quality is high, and results are interpreted and used effectively. To catch Kerviel, SocGen needed a system that could track all finalized, modified, pending, and canceled transactions as well as deals with deferred starting dates and internal counterparties. It also needed to compare "off-market" to market prices, monitor external brokers' fees, reconcile transactions with internal counterparties, and confirm all bookings for intramonthly provisions.

Front and Back Offices Must Be Separated

Neither Barings nor SocGen put Chinese walls between their front office (i.e., traders who transact business with customers) and the middle/back offices (i.e., individuals responsible for booking, clearing, monitoring, and controlling transactions). As a result, Leeson and Kerviel were able to violate their internal trading limits. Barings also made the egregious mistake of consciously empowering Leeson with both front office *and* back office supervisory responsibilities at BFS, thereby enabling him to trade and also to measure his own profitability.

Kerviel accomplished the same results as Leeson by obtaining back office usernames, passwords, and control codes. He was also an artist at financial double-talk. Often back office and middle office staff did not understand Kerviel's explanations of accounting irregularities. Because of the culture and power structure within SocGen, staff did not go out of their way to verify the truth of Kerviel's statements, challenge him, or transmit their concerns to supervisors.

Proprietary and Agency Business Must Be Separated

The risk management systems at Barings failed to separate funds invested on behalf of the bank (proprietary business) from funds invested for customers (agency business). Leeson put Barings' equity at enormous risk, but London headquarters was convinced that any net positions Barings had with the Asian exchanges were solely on behalf of customers. Barings (London) ended up transferring more than £300 million to Leeson in Singapore without ever checking to see if there were sufficient funds in customers' accounts, if customers had exceeded their exposure limits, or if the customers were creditworthy for this volume of business (Bank of England, 1995, Sections 13.22–13.23).

Gross* and *Net Positions Must Be Monitored and Controlled

Barings and SocGen focused their controls on traders' net positions, almost to the exclusion of gross positions. Emphasizing a trader's net position is logical because long and short positions in identical contracts should offset each other, rendering them harmless. But focusing exclusively on net exposures assumes that a trader's reported positions are accurate and up-to-date, which was not the case at Barings and SocGen. The banks discovered too late that Leeson and Kerviel were placing colossal one-sided bets that were far in excess of these financial institutions' equity bases.

The control system at SocGen also made the mistake of tracking aggregate (i.e., divisional) trading activities rather than those of any individual trader. As a result, Kerviel's excessive transactions and margin payments were blended with the transactions of other traders, which made his trades, positions, and cash flows appear innocuous. Only afterward did supervisors learn that, between April 2007 and November 2007, Kerviel was responsible for 25 to 60 percent of the net margin payments made by SocGen's Equity Derivatives Division.

Compensation Incentives and Promotion Criteria Must Be Scrutinized

A well-designed compensation plan should reflect the value that employees add to a company's success and provide incentives for employees to work in ways that are consistent with company goals. If an employee is paid a fixed annual salary, then business risks are borne almost exclusively by the company, but if part of an employee's compensation is tied to performance, then a portion of these business risks are (presumably) transferred to the employee.

Many hedge funds operate under a "2 and 20" compensation policy, earning a 2 percent fee for funds under management and 20 percent of any profits after a high-water mark is reached. Profits typically are recorded at year-end and are based on both realized and unrealized gains. Bonuses are also based on realized and unrealized gains, but if these profits are never realized, this portion of the bonuses is not returned to investors.

Leeson earned bonuses of £35,746 in 1992 and £130,000 in 1993, and he was scheduled to earn £450,000 in 1994, before Barings discovered his deception. Hunter earned a $75 million bonus in 2005 alone, which was far above his annual salary. In fiscal year 2006, Kerviel earned a €60,000 bonus on top of his €74,000 salary, and he expected to earn a €300,000 bonus from his trading profits in 2007, before SocGen discovered that he had been cooking the books.

Bonus compensation schemes provide traders with little or no incentive to lower risks. In fact, they may give heavy inducements to increase them. The two major forces restraining traders from increasing their risks to harmful levels are the limits imposed by a company's risk management system and the ethical character (i.e., moral standards, honesty, and integrity) of the traders. It is not an exaggeration to say that most traders, in the normal course of their jobs, spot holes in their companies' risk management systems that could be exploited for profit, personal gain, and/or fame. Most of them choose not to do so because of the people they are and not because of their chances of being caught. In the end, no risk management system can totally prevent abuses if traders and their managers do not have their own high standards of ethical behavior.

Even though compensation is a significant incentive influencing traders' and fund managers' behavior, it is not always the most important. Neither of our rogue traders (Leeson nor Kerviel) used bank resources to directly increase their personal wealth (i.e., there was no strong evidence of embezzlement). Clearly, these traders were motivated by the potential for bonus compensation, but equally, if not more, important were the nonpecuniary rewards that came from being admired as *top traders*.[13] Leeson and Kerviel came from middle-class families. Neither trader had ties to the nobility, high-ranking corporate leaders, or distinguished politicians, nor did they have degrees from *elite* universities. Promotion to the jobs they had was a wonder in itself.

Better Ways Are Needed to Distinguish Skill from Luck

If derivative prices reflect all the information that is currently available to the market, then traders can consistently earn above-average profits only if they possess insider information or superior skills, take above-average risks, or are lucky. Determining the basis for success is important because successful traders and trading strategies are rewarded with expanded trading authority.

Jérôme Kerviel's first exposure to "trader stardom" came in 2005 when he sold short Allianz (a German insurance company) shares just before terrorists bombed the London Underground on July 7. His gains (€500,000) were the result of an event that he had no way of anticipating and that was not part of an investment "strategy." Similar good fortune struck Brian Hunter in 2005 when he earned more than $1.26 billion for Amaranth and was handsomely rewarded with a $75 million bonus and numerous other perks. Hunter's strategy was to purchase in early 2005 (what he thought were) underpriced call options. By June 2005, it looked as if his expectations were completely wrong. Energy prices remained tepid, and his options were significantly out-of-the-money. Nevertheless, with Forest Gump–like luck, energy prices spiked when two unexpected hurricanes (Katrina and Rita) wiped out a substantial portion of the Gulf of Mexico's production capacity in August and September. Hunter was rewarded handsomely for his profits and *keen* foresight when, in fact, his success seems to have been due almost exclusively to two storms that he had no way of predicting in early 2005.

A Culture of Ethical Behavior Needs to Be Developed

The derivative disasters at Barings and SocGen provided evidence that combining dishonest traders with weak risk management systems is a formula for disaster. They also highlighted the need for companies to develop a culture of ethical behavior that engenders in employees a sense of accountability, shared values, and mutual respect. By outward appearances, Leeson and Kerviel were conscientious employees who took few holidays and worked well into the night. Who would have guessed they were actually spending their time falsifying trades, hacking into computer systems, and spinning webs of deception? Ideally, individuals who are given increased trading authority would be worthy of the trust placed in them. When Leeson and Kerviel were promoted to traders from their back office and middle office positions, Barings and SocGen must have felt that they had reasonable, firsthand information about the honesty, integrity, and work ethic of these men.

"Trust but verify" should be the motto of every risk management system; unfortunately, it is not. Barings seemed to base its system on the age-old city of London maxim: "My word is my bond." Expecting its employees to *do the right thing*, Barings was crippled and then finished off by a lethal combination of an unscrupulous trader who did not deserve the trust that was placed in him and a horrifically disorganized risk management system that put almost no barriers in his way. Similarly, Jérôme Kerviel's managers at SocGen relied on trust more than meaningful oversight.

The financial catastrophes surrounding Leeson and Kerviel point to classic problems of asymmetric information (i.e., principal-agency, adverse selection, and moral hazard). Principal-agency problems emerged from the risks these traders took, which were not in the best interest of the banks. Even though they acted to increase shareholder value, their exposures were unauthorized, unreasonable, and ended up costing Barings and SocGen billions in losses. These financial institutions were also victims of adverse selection problems because Leeson and Kerviel (i.e., the agents) had informational powers over their employers, which were not revealed during the hiring process. Both men were attracted by the opportunity to earn bonus compensation that was far above the pay scales of their former jobs, but equally (if not more) important was the opportunity to become envied "star traders" who could stand shoulder-to-shoulder with elite bank managers. If they succeeded, their dreams would be fulfilled. But if they failed, their banks would bear the financial burdens, and these traders would simply fade back into employment positions they expected to occupy anyway.

Moral hazard problems arose because these financial institutions were unable or unwilling to pay the price needed to accurately measure, monitor, and control their employees' risk-adjusted performances. As a result, Leeson and Kerviel took risks they would not have taken if Barings and SocGen had better risk management systems. In a way, these traders' insurance policies were embedded in their intimate knowledge of Barings' and SocGen's back office information flows and risk management systems—information they gained while working in these functions prior to becoming traders.

Risk Management Systems Are Only as Strong as Their Weakest Risk Managers

Even the best-designed risk management systems are going to produce disappointing results if managers do not understand the significance of and interrelationships among credit risk, market risk, and liquidity risk. Each of these risks has its own independent effect on a company's cash flows, income statement, and balance sheet, but they are also highly interrelated. Like deer crossing a road, one type of risk is rarely seen alone, especially in times of crisis. Equally important are back office and middle office staff, who work to the job and not to the hour or to the letter of their job descriptions. Rules of behavior and job descriptions cannot cover every eventuality. Therefore, companies need employees who insist on clear answers to accounting questions, take serious concerns to their supervisors, and use healthy doses of common sense to perform daily tasks.

MG incurred significant losses in 1993 because its German-based supervisors had a poor understanding of the cash flow risks associated with the stack-and-roll hedges used by American-based MGRM. Supervisors did not realize that the success or failure of these hedges cannot be evaluated by interim cash flows. In fact, midstream assessments of profitability are likely to give precisely the wrong impression (Marthinsen, 2009, p 104–110). MGRM's German managers also failed to grasp that part of the problem was an accounting illusion, due to the difference between German rules, which required German-based companies to value their positions using the lower-of-cost-or-market (LCM) method, and U.S. rules, which permitted hedge accounting.

Amaranth appeared to have a world-class risk management system because its risk managers created daily reams of position statements, profit and loss reports, as well as analyses measuring value at risk, premium at risk, stress tests, leverage, concentrations, industry exposures, and portfolio sensitivities to changes in price, time, interest rates, and volatility. Furthermore, Amaranth distinguished itself by having a designated risk manager at every trading desk. But even the most sophisticated risk management system in the world (Amaranth's was not) would be of little value if risk managers failed to use it to control the exposures of traders, like Brian Hunter.

The level of mismanagement at Barings may go down in history as a low-water mark from which to measure all future supervisory blunders. When the dust cleared, it was obvious that Barings' risk management system was weak, but it was equally obvious that the negligence and blasé attitude of Leeson's supervisors and Barings' crew of risk managers were fully responsible for allowing Leeson to operate as long as he did. Internal and external audits of BFS were slipshod. No one in the Barings organization was ever quite certain whether Leeson was in charge of settlements, compliance, arbitrage, proprietary trading, or executing customers' orders. Through their acts of omission, rather than commission, Leeson's supervisors allowed him to violate trading limits, falsify reports, misstate profits, record nonexistent transactions, make unauthorized trades, and fabricate accounting entries. Barings' managers had numerous opportunities to uncover Leeson's deceptive trades and practices long before the bank's failure, but they never did.

The quality of managerial oversight at SocGen was no better. Controls at the bank were fragmented, often being split among different parts of the same function or among different functions, thereby, inhibiting a healthy overview of firm-wide risks. Between late July 2006 and September 2007, SocGen's supervisors ignored 64 warnings from the bank's risk department about activities that had either direct or indirect links to Kerviel's fraudulent activities (including two loans, each amounting to €500 million, to support his transactions). Managers did little or nothing when auditors found SocGen's internal controls to be weak and lacking. SocGen's chief executive officer, Daniel Bouton, never investigated with sufficient depth the French banking commission's warnings and alarms that better control systems were needed at SocGen, particularly in the equity derivatives area. Finally, complaints, warnings, and letters of inquiry from trading floor inspectors and exchanges (e.g., EUREX) about Kerviel's suspicious trading activities were disregarded by SocGen managers.

BROADER IMPLICATIONS OF DERIVATIVE SCANDALS AND DISASTERS

As horrific as our five derivative scandals and disasters were, none of the companies missed a margin call or failed to make a debt payment. Therefore, the solvency of creditors, exchanges, and counterparties was never seriously threatened. Neither central banks nor governments bailed out these financially distraught companies, and the disasters turned out to be implosions, rather than explosions, with no significant spillover effects (i.e., contagion) to broader financial markets or real sectors.[14] In fact, some companies benefited handsomely from these failures (e.g., those that purchased the positions of failing companies and then sold them shortly thereafter at considerable profits).

Even though federal assistance was not needed, bailouts were clearly an option, as the 2008–2009 financial and economic crisis proved. The role of central banks as ultimate sources of liquidity and governments as caretakers of last resort leads us to the issue of moral hazard. Federal regulators and central banks must walk a tightrope between their responsibility to ensure healthy financial systems by preventing systemic risks and their obligation to remain detached from idiosyncratic risks that individual companies choose to take and for which they are rewarded. Companies have different appetites for risk and, when they get it right, are the beneficiaries of their successes. If these companies are bailed out by central banks or governments whenever they are threatened with failure, then others might also be encouraged to take imprudent risks. Such behavior could have deleterious effects on the growth and development of our money and capital markets.

Perhaps the world was lucky to have escaped from these financial crises with relatively little or no significant harm. To be sure, these scandals and disasters may have made the markets stronger and wiser. Despite the incredibly large losses suffered by banks, hedge funds, and companies, our global regulatory systems and exchanges seem to have functioned rather effectively. One of the major reasons was because, as colossal as these losses were, they were relatively small in comparison to the global money and capital markets. Therefore, efforts to prevent future catastrophes must make sure that neither the growth of our financial markets nor the development of international living standards is inhibited. Nevertheless, there still remain major reasons for concern, and, among them is whether our financial markets would be able to withstand multiple simultaneous failures the size of LTCM, Amaranth, and SocGen. As banks, insurance companies, and many other financial institutions failed or were threatened with failure in 2008 and 2009, the threat they pose to the well-being of us all was clearly apparent.

MGRM was saved by a consortium of its largest shareholders; Barings was sold to ING, a Dutch bank; LTCM was recapitalized by a consortium of 14 banks and brokerage houses; Amaranth's positions were purchased by Citadel and JPM; and finally, as extraordinarily large as they were, the losses at SocGen were not enough to bankrupt the bank. In February 2008, SocGen launched a successful €5.5 billion (nearly $8 billion) rights issue to recapitalize itself and to restore its top-tier ranking.

CONCLUSION

Risk management systems must grow in size and sophistication with the businesses they monitor and control. The need for better self-policing is obvious, as is the need for more effective liquidity management, valuation models, and OTC settlement procedures. Traders' activities must be scrutinized, with limits put on their gross and net positions, both on-balance sheet and off-balance sheet. Firm-wide risks and cross-functional sharing of quantitative and qualitative valuation information also need to be part of the risk management process (Senior Supervisors Group, 2008). Moreover, counterparty risk management needs to be strengthened so that companies are protected from direct credit risks and the indirect market risks of multiple counterparty failures.

Responsibility for controlling risks on a day-to-day basis should be integrated into operating business lines so that performance evaluations and compensation rewards encourage traders and fund managers to consider *risk-adjusted* returns. Top-line goals, such as contract growth and quality of execution, need to be tempered with better risk management practices. Assessments of firm-wide exposures should be performed by individuals who are independent and have considerable organizational authority. Risk managers should have years of experience and understand how to balance objective risk criteria (e.g., VaR) with qualitative inputs from a broad cross-section of the firm. These responsibilities cannot be the job of individuals who are perceived within the organization to be cost centers that serve only to reduce profitability.

Open and constructive dialogs within a company might be triggered by setting VaR thresholds at lower limits so they are violated more often and prompt managerial inquiries into risk forecasts. Other suggestions are to perform periodic, firm-wide stress tests that identify common risk drivers and consider worst-case scenarios. Because sunlight is the best of all disinfectants, a key ingredient in any solution is transparency. By clearly communicating what is a stake, financial transparency can help to ensure that wise business decisions are made so that derivative scandals and disasters can be avoided in the future.

ACKNOWLEDGMENTS

I am very grateful to Amar Sahay for his thoughtful comments, suggestions, and feedback on drafts of this paper.

ENDNOTES

1. All of these scandals and disasters, except Société Générale, are explained in detail in Marthinsen (2009). Details of Société Générale's disaster can be found in Société Générale (2008).
2. During the early 1990s, about 90 percent of the oil futures contracts traded on NYMEX expired within four months, which highlights the considerable value that MGRM brought to the energy markets.
3. MGRM's stack-and-roll hedges involved purchasing enough short-dated energy futures contracts (i.e., futures contracts with maturities less than one year) to neutralize the expected profits/losses of its long-dated forward contracts.
4. Basis is the difference between the spot price and forward (or futures) price of an underlier.
5. Unbalanced cash flows occurred because exchanged-traded stack-and-roll hedges were marked to market on a daily basis. In addition, they required MGRM to constantly roll over its stack of short-dated futures contracts (i.e., its hedges) while the long-dated forward contracts matured at a relative snail's pace.

6. In 1998, LTCM's leverage ratio rose from 30:1 to 100:1. This increase was mainly the result of a precipitous decline in the shareholders' equity (due to losses) and a more sluggish decline in asset values. See Lowenstein (2000, p. 78).
7. See President's Working Group on Financial Markets (1999, pp. 14 and C-13).
8. Amaranth's large share of the open interest on derivative exchanges, such as the New York Mercantile Exchange and IntercontinentalExchange, is indicative of the fund's considerable market presence but not necessarily its market power because there had to be counterparties on the other side of Amaranth's positions. To date, no single counterparty or small group of counterparties has been reported to have had as large a share as Amaranth. See United States Senate Permanent Subcommittee on Investigations Committee on Homeland Security and Governmental Affairs, Excessive Speculation in the Natural Gas Market and Appendix, Washington D.C. Government Printing Office, 25 June (2007). Staff Report—Excessive Speculation in the Natural Gas Market; available at http://hsgac.senate.gov/public/_files/REPORTExcessiveSpeculationintheNaturalGasMarket0.pdf. Appendix—Excessive Speculation in the Natural Gas Market; available at http://levin.senate.gov/newsroom/supporting/2007/PSI.Amaranthappendix.062507.pdf. Accessed 22 February 2009.
9. After Amaranth's failure in 2007, the Commodity Futures Trading Commission (CFTC) accused the fund of intentionally and unlawfully attempting to manipulate NYNEX natural gas futures prices on two separate contract expiration days. The Federal Energy Regulatory Commission also accused Amaranth of manipulating (i.e., not just attempting to manipulate) physical natural gas prices on three separate occasions and proposed penalties equaling $291 million. These price manipulation charges and the alleged strategies behind them are explained in Marthinsen (2009).
10. Leeson was supposed to be arbitraging futures contracts (e.g., Nikkei 225, Japanese government bonds, and euroyen futures contracts) between Asian exchanges, and Jérôme Kerviel was supposed to be arbitraging Turbo warrants (e.g., covered warrants with knock-out options) between European futures exchanges and the forward markets.
11. In late 2007, Amaranth filed suit for more than $1 billion in damages against its clearing agent, JPM. The fund claimed that JPM abused its position by preventing Amaranth from negotiating a better buyout deal with Goldman Sachs. Amaranth also claimed that its $2.5 billion concession payment was unwarranted and that JPM caused additional harm.
12. Jérôme Kerviel's reported profits increased by 514 percent (from €7 million to €43 million) between 2006 and 2007 without raising the suspicions of his supervisors. The weaknesses of SocGen's internal control systems, organizational structures, and governance policies, as well as remediation measures taken to address these problems, can be found in PricewaterhouseCoopers (2008).
13. In 2007, Kerviel reported profits on his proprietary trades of €25 million, placing him among the top 11 percent of SocGen's 143 arbitrage traders. Corrected figures later revealed that he was nothing more than a mediocre trader.
14. The "subprime crisis," which began in 2007 and spread during 2008 and 2009 to the financial and real sectors of the United States and the rest of the world, was not included in this chapter. Even though the misuse of financial derivatives (e.g., credit default swaps) played an important role, the causes of this debacle were much more widespread, including reckless securitization, regulatory errors, lost trust, crushed confidence, monetary policy mistakes that led to speculative bubbles, systemic risk, and contagion. The U.S. Federal Reserve's involvement in the 1998 rescue of LTCM may have portended its even large role in 2008 and 2009, when the Fed rescued many U.S. financial institutions from dire liquidity shortages and mounting insolvency threats. In 1998, the Fed applied considerable jawboning pressure (moral suasion) to facilitate LTCM's recapitalization by a consortium of private financial institutions. Direct loans

to LTCM were considered problematic because the central bank had no jurisdictional powers over hedge funds and because of moral hazard concerns. By contrast, the Fed's involvement during the 2008–2009 subprime, financial, and economic crisis was much more aggressive, involving loans and guarantees that totaled hundreds of billions of dollars.

REFERENCES

Bank of England, Board of Banking Supervision. 2005. Report of the Board of Banking Supervision Inquiry into the Circumstances of the Collapse of Barings London (ordered by the House of Commons) (July).

Brown, S. J., and O. W. Steenbeek. 2001. "Doubling: Nick Leeson's Trading Strategy," *Pacific-Basin Finance Journal* 9: 83–99.

Lowenstein, R. 2000. *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. New York: Random House.

MacKenzie, D. 2003. "Long-Term Capital Management and the Sociology of Arbitrage," *Economy and Society* 32, no. 3 (August): 349–380.

Marthinsen, J. 2009. *Risk Takers: Uses and Abuses of Financial Derivatives*, 2nd ed. Boston: Prentice Hall.

President's Working Group on Financial Markets. 1999. *Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management: Report of the President's Working Group on Financial Markets*. Washington, DC: Department of Treasury, April 28.

PricewaterhouseCoopers. 2008."Société Générale: Summary of PwC Diagnostic Review and Analysis of Action Plan" (May 23); available at: www.efinancialnews.com/downloadfiles/ 2008/05/2450756079.pdf. Accessed 22 February 2009.

Scholes, M. S. 2000. "The Near Crash of 1998: Crisis and Risk Management." *American Economic Review* 90, no. 2 (May): 17–21.

Senior Supervisors Group. 2008. "Observations on Risk Management Practices during the Recent Market Turbulence" (March 6); available at: www.newyorkfed.org/newsevents/news/ banking/2008/SSG_Risk_Mgt_doc_final.pdf. Accessed 22 February 2009.

Société Générale. 2008. "General Inspection Department, Mission Green: Summary Report" (May 20); available at: www.iht.com/pdfs/business/sgreport3.pdf. Accessed 22 February 2009.

Swan, E. J. 2000. *Building the Global Market: A 4000 Year History of Derivatives*. London: Kluwer Law International.

Till, H. 2006. "EDHEC Comments on the Amaranth Case: Early Lessons from the Debacle." Lille, France: EDHEC Risk and Asset Management Research Centre and Principal and Premia Capital Management, LLC, EDHEC Business School, 2006; available at www.edhec-risk.com/features/RISKArticle.2006-10-02.0711/attachments/EDHEC%20Comments%20on%20Amaranth%20Case.pdf. Accessed 22 February 2009.

United States Senate Permanent Subcommittee on Investigations Committee on Homeland Security and Governmental Affairs. 2007. "Excessive Speculation in the Natural Gas Market and Appendix." Washington, DC: Government Printing Office (June 25), p 1–135. Available at http://hsgac.senate.gov/public/_files/REPORTExcessiveSpeculationintheNaturalGasMarket0.pdf.

United States Senate Permanent Subcommittee on Investigations Committee on Homeland Security and Governmental Affairs. 2007. "Excessive Speculation in the Natural Gas Market and Appendix." Washington, DC: Government Printing Office (June 25), p 136–V38. Available at http://levin.senate.gov/newsroom/supporting/2007/PSI.Amaranthappendix.062507.pdf.

SUGGESTED FURTHER READING

Collins, D. P. 2007. "Manipulating a Hedge Fund Blow-up," *Futures* 36, no. 11 (September): 66–68.

Culp, C., and M. Miller. 1999. *Corporate Hedging in Theory and Practice: Lessons from Metallgesellschaft*. London: Risk Publications.

Dunbar, N.. 2000. *Inventing Money: The Story of Long-Term Capital Management and the Legends behind It*. New York: John Wiley & Sons.

Edwards, F. S., and M. S. Canter. 1995. "The Collapse of Metallgesellschaft: Unhedgable Risks, Poor Hedging, or Just Bad Luck?" *Journal of the Futures Market* 15: 211–264.

Fay, S.. 1997. *The Collapse of Barings*. New York: W. W. Norton.

Kambhu, J., T. Schuermann, and K. J. Stiroh. 2007. "Hedge Funds, Financial Intermediation, and Systemic Risk," *Federal Reserve Bank of New York Economic Policy Review* (December): 1–18.

Krapels, E. 2001. "Re-examining the Metallgesellschaft Affair and Its Implications for Oil Traders," *Oil & Gas Journal* (March 26): 70–77.

Leeson, N., with E. Whitley. 1996. *Rogue Trader: How I Brought Down Barings Bank and Shook the Financial World*. Boston: Little, Brown.

Lim Choo San, M., and N. Ng Kuang Tan. 1995. "Barings Futures (Singapore) Pte Ltd: Investigation Pursuant to Section 231 of the Companies Act (Chapter 50): The Report of the Inspectors Appointed by the Minister for Finance." Singapore: Singapore Ministry of Finance, pp. xi, 183.

Mello, A. S., J. E. Parsons, C. L. Culp, and M. H. Miller. 1995. "Maturity Structure of a Hedge Matters: Lessons from the Metallgesellschaft Debacle," *Journal of Applied Corporate Finance* 8, no. 1 (Spring): 106.

Muehring, K. 1996. "John Meriwether by the Numbers," *Institutional Investor* 30, no. 11 (November): 68–81.

Pirrong, S. C.. 1997. "Metallgesellschaft: A Prudent Hedger Ruined, or a Wildcatter on NYMEX?" *Journal of Futures Markets* 17, no. 5: 543–578.

Poole, W. 2008. "Market Bailouts and the 'Fed Put,'" *Federal Reserve Bank of St. Louis Review* (March/April): 65–73.

Société Générale. 2008. *Progress Report of the Special Committee of the Board of Directors of Société Générale* (February 20); available at: www.sp.socgen.com/sdp/sdp.nsf/V3ID/6D44E7AEF3D68993C12573F700567904/$file/comiteSpecialFevrier08gb.pdf. Accessed 22 February 2009.

Taleb, N. N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Wollert-Elmendorff Deutsche Industrie-Treuhand GmbH und C&L Treuarbeit Deutsche Revision. 1995. *Special Audit of Metallgesellschaft Aktiengesellschaft*. C&L Treuarbeit Deutsche Revision/Wollert-Elmendorff Deutsche Industrie-Treuhand, 1995, *Bericht über die Sonderprüfung nach § 142 Abs. 1 AktG bei der Metallgesellschaft AG, Frankfurt am Main, gemä Beschlu der auerordentlichen Hauptversammlung am 24. Februar 1994*.

ABOUT THE AUTHOR

John E. Marthinsen is professor of economics and international business at Babson College in Babson Park, Massachusetts, where he holds the Distinguished Chair in Swiss Economics of the Glavin Center for Global Entrepreneurial Leadership. Dr. Marthinsen received his BA from Lycoming College and both his MA and PhD from the University of Connecticut. He has extensive consulting experience, working for both domestic and international companies as well as for the U.S. government. Dr. Marthinsen's primary research interests are in the areas of corporate finance and international financial markets. He is an award-winning teacher and the author of many articles and books. Among Dr. Marthinsen's most recent books are *Risk Takers: Uses and Abuses of Financial Derivatives*, 2nd ed. and *Managing in a Global Economy: Demystifying International Macroeconomics*. Dr. Marthinsen served on the United Nations Association's Economic Policy Council and lectured at the University of Bern and Basel in Switzerland and the University of Nürenberg, Germany. He is a member of the board of directors of Givaudan SA, a Swiss-based flavors and fragrances company.

PART IV

Pricing of Derivatives: Essential Concepts

Virtually all pricing strategies for financial derivatives take account of the zero-sum nature of derivatives contracts and exploit the concept of freedom for arbitrage in pricing. Derivatives contracts are zero sum in the sense that the buyer's gains equal the seller's losses, so adding all gains and losses gives a total equal to zero. (This ignores transaction costs and other related market frictions that actually make derivative contracts negative sum by the amount of those frictions.) This is an important idea, because two parties, a buyer and seller, both of whom contract in their own interest and are aware of the zero-sum nature of the market, consummate a transaction. Given this awareness and a commitment to their self-interest, neither party would agree to a contract that gave the other a certain profit after taking into account the cost of the invested funds. This is the no-arbitrage principle, which states that prices in the derivatives market must be such that there can be no certain profit without investment, because to provide one party with this kind of arbitrage profit would mean that the other party was accepting a certain loss of a magnitude equal to the arbitrage gain. Part Four introduces the essential concepts of derivatives pricing by exploiting this no-arbitrage principle. (Part Five extends the ideas developed here.)

This idea of arbitrage-free pricing is the topic of Chapter 24, "No-Arbitrage Pricing," by Robert A. Strong. Strong explains how this principle is used in derivatives pricing by giving concrete examples. He further extends the analysis by showing that the no-arbitrage condition requires that two portfolios of derivatives securities with exactly the same payoffs in all situations must have the same price to preclude arbitrage.

David Dubofsky applies the no-arbitrage principle in Chapter 25, "The Pricing of Forward and Futures Contracts." Spot prices and prices for future delivery are mediated by the cost of carry, the cost of acquiring a good and storing it to deliver against a forward or futures contract initiated at an earlier contract date. As Dubofsky shows, the spot price of a good, its forward or futures price, and the cost of storing it from the present to the future delivery date must form an integrated system of prices that conform to the cost-of-carry relationship and thus preclude arbitrage. The application of this concept becomes more complicated, as Dubofsky explains, when issues of storage and wastage along with potential shortages enter the analysis.

A. G. Malliaris explains the pricing of option contracts in Chapter 26, "The Black-Scholes Option Pricing Model." The great achievement of Black and Scholes was to show how to apply no-arbitrage pricing within a framework of continuous time mathematics that allows one to compute exact theoretical option prices. As time has proven, this model has tremendous application in actual market.

Soon after the publication of the Black-Scholes model and its digestion by finance scholars, researchers began to find many more applications for models in the spirit of Black-Scholes—models that exploit continuous time mathematics and yield pricing models for securities that are closed form. António Câmara explores these extensions in Chapter 27, “The Black-Scholes Legacy: Closed-Form Option Pricing Models.” Câmara shows how the Black-Scholes spirit was extended to more complex kinds of financial derivatives. Whereas the Black-Scholes model focused on an option that depends on one lognormally distributed variable, other options depend on more such variables or on variables that follow other stochastic processes.

Gerald D. Gay and Anand Venkateswaran apply the no-arbitrage pricing principle to Chapter 28, “The Pricing and Valuation of Swaps.” Pricing of swaps proceeds by carefully applying the no-arbitrage principle and the concept of the time value of concepts, based on the idea that the value of the exchanged cash flows must have equal present values. The basic types of swaps are interest rate and foreign exchange swaps. Both involve the term structure of interest rates, so an essential part of pricing swaps requires an understanding of the term structure, which Gay and Venkateswaran develop.

CHAPTER 24

No-Arbitrage Pricing

ROBERT A. STRONG, CFA
Foundation Professor of Investment Education and Professor of Finance,
University of Maine

Finance sometimes is called the study of arbitrage. *Arbitrage* is the existence of a riskless profit. Risk and expected return are generally proportional, so we would not expect to find riskless profit opportunities very often, and if they do appear, they should quickly disappear as traders take advantage of them. This is exactly what happens in practice.

It is important to note that finance theory does not say that arbitrage will never appear. Rather, theory states that arbitrage opportunities will be short-lived: The market will act quickly to eliminate the arbitrage and bring prices back into equilibrium.

FREE LUNCHES

Suppose you are in a European train station and see these exchange rates posted for euros, Australian dollars, and U.S. dollars:

€1.00 = \$1.1517
AUD1.00 = \$0.8306
€1.00 = AUD1.4022

This set of exchange rates provides an arbitrage opportunity, as Exhibit 24.1 shows.

Sometimes the apparent mispricing is too small to be worth exploiting. You can find pennies on the ground in any parking lot. We have all seen them, and many of us just let them stay there. They are not worth the trouble to pick up. A \$20 bill, however, would quickly be snatched up.

Other times an apparent arbitrage opportunity is out of reach because of some impediment to free trade or some other restriction. You might pay 85 cents to enter a subway station and then notice two quarters on the floor just outside the gate you entered. To go get them you would have to go back through the turnstile and then deposit another 85-cent token to get back in. Fifty cents lying on the ground is textbook arbitrage, but not if you have to pay 85 cents to get it. You might move quickly into action if the money were a \$20 bill rather than a couple of coins, but even this is not riskless. There are lots of people in a subway station, and you run

Exhibit 24.1 Foreign Currency Arbitrage

Buy €10,000 with U.S. dollars: cost = €10,000 × \$1.1517/€ = \$**11,517.00**
Exchange the euros for Australian dollars: €10,000 × AUD1.4022/€ = AUD14,022.00
Exchange the Australian dollars for U.S. dollars: AUD14,022.00 × \$0.8306/€ = \$**11,646.67**
This is an arbitrage profit of \$11,646.67 − \$11,517.00 = \$**129.67.**

the risk of having someone else pick up the bill by the time you get to it, in which case your "trading fee" is simply lost money.

Modern option pricing techniques are based on arbitrage principles. Certain packages of securities are equivalent to other packages. In a well-functioning marketplace, equivalent assets should sell for the same price. Given the required information, we can solve for what an option price must be for arbitrage to be absent. The classic study (Stoll, 1969) of arbitrage in option pricing gave birth to the term *put/call parity*, the subject of the next section. This relationship shows that for a given underlying asset the call price, put price, stock price, and interest rate form an interrelated complex, and that given three of the values, you can solve for the fourth.

THEORY OF PUT/CALL PARITY

As Exhibit 24.2 shows, the characteristic shape of the profit/loss diagram[1] for a covered call position is the same as that of a short put. What happens if you combine a covered call with a *long* put? The diagram for a long put is obtained by rotating the short put about the horizontal axis: A long put is exactly the opposite of a short put. With European options and a non–dividend paying stock, an investor who combines a long stock position with a short at-the-money call and a long at-the-money put has a *riskless position*. (See Exhibits 24.3 and 24.4.) The combination of a short put, a short stock position, and a long call also yields a riskless position, and riskless investments should earn the riskless rate of interest, if the riskless position requires you to advance funds.[2]

Suppose an investor borrows money to buy stock and simultaneously writes a call and buys a put, with both options at-the-money. The investor then holds this position until option expiration. According to Exhibit 24.4, this results in a perfect hedge because no matter what happens, the future value of the position is fixed. Because the options are at-the-money, the stock price and strike price are equal. If the stock rises, it will be called away at the strike price. If it falls, you can sell it

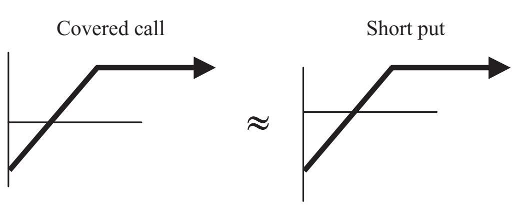


Exhibit 24.2 Covered Calls and Short Puts

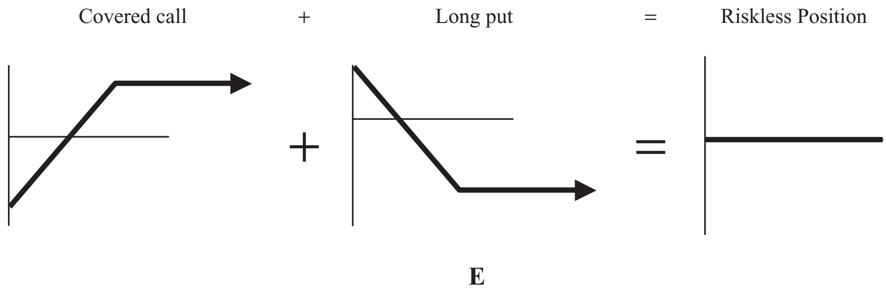


Exhibit 24.3 Covered Calls and Long Puts

at the strike price. Your stock and options portfolio will be worth the stock price regardless of what the stock itself does.

Because this is the only possible outcome, it is riskless, and, in theory at least, a bank should be willing to lend money for this position at r, the riskless rate of interest for the period until expiration. If an investor can establish these three positions and make a profit, arbitrage is present. Arbitrage profits should equal zero, so the market will adjust such that

$$S - S + C - P - \frac{Sr}{(1+r)} = 0 \tag{24.1}$$

or

$$C - P - \frac{Sr}{(1+r)} = 0 \tag{24.1a}$$

where C = call premium
P = put premium
S = stock price
r = riskless interest rate

The equation comes from this logic. After establishing the three positions, there are one cash *inflow* (from writing the call) and two cash *outflows* (paying for the put and paying the interest on the bank loan). The principal of the loan (S) comes in, but I immediately spend it ($-S$) to buy the stock. The interest on the bank loan is paid in the future: It needs to be discounted to a present value. That is why the interest charge (Sr) is divided by the quantity $(1 + r)$.

Exhibit 24.4 Stock Price at Option Expiration

	0	**$25**	**$50**	**$75**
Buy stock @ $50	0	25	50	75
Write $50 call	0	0	0	(25)
Buy $50 put	50	25	0	0
Total	$50	$50	$50	$50

We can rearrange the equation as shown in Equation 24.2.

$$C - P = \frac{Sr}{(1 + r)} \tag{24.2}$$

Dividing both sides of the equation by the price of the stock (S), we get

$$\frac{C}{S} - \frac{P}{S} = \frac{r}{(1 + r)} \approx r \tag{24.3}$$

or

$$\frac{C - P}{S} = \frac{r}{(1 + r)} \approx r \tag{24.3a}$$

The quantity $r/(1 + r)$ is approximately equal to r. Suppose, for instance, the interest rate is 5 percent. Five percent divided by 1.05 gives 4.76 percent. On a \$25 stock, the difference between the one-year call and put premium, according to Equation 24.3a, should be (4.76%) × (\$25) = \$1.19. On a \$100 stock, however, the difference would be *four times* as much: (4.76%) × (\$100) = \$4.76.

The point is that relative put and call prices differ by about the riskless rate of interest. In other words, the call premium should exceed the put premium, and the difference will be greater as the price of the stock goes up, as interest rates rise, or as the time to option expiration lengthens. A simple example will show the implications of this. First let us expand the list of variables:

C = call premium	K = option strike price
P = put premium	r = riskless interest rate
S_0 = current stock price	t = time until option expiration
S_1 = stock price at option expiration	

Suppose we do as before: Write the call, buy the put (with the same strike price as the call), and buy stock, but instead of borrowing the current stock value, we borrow the *present value* of the *strike price* of the options, discounted from the option expiration date. If the options are at-the-money, the stock price is equal to the option strike price. It is necessary to discount the strike price, because this amount is paid in the future, and dollars today are not the same as dollars tomorrow. This yields a profit/loss contingency table for the combined positions as Exhibit 24.5 shows.

Regardless of whether the stock price at option expiration is above or below the exercise price, the net value of the combined positions is zero: $C - P - S_0 + \frac{K}{(1+r)^t} = 0$. Rearranging terms, we have the classic *put/call parity* relationship:

$$C - P = S_0 - \frac{K}{(1 + r)^t} \tag{24.4}$$

Exhibit 24.5 shows that call prices, put prices, the stock price, and the riskless interest rate form an interrelated securities complex. If you know the value of

Exhibit 24.5 Put-Call Parity Arbitrage Table

		Value at Option Expiration		
Activity	**Cash Flow**	**If $S_1 < K$**	**If $S_1 > K$**	**If $S_1 = K$**
Write call	$+C$	0	$K - S_1$	0
+ Buy stock	$-S_0$	S_1	S_1	S_1
+ Buy put	$-P$	$K - S_1$	0	0
+ Borrow	$K/(1 + \mathrm{r})^t$	$-K$	$-K$	$-K = -S_1$
= Sum	$C - P - S_0 + K/(1 + r)^T$	0	0	0

three of these components, you can solve for the equilibrium value of the fourth. The relationship assumes that the options can be exercised only at expiration and that the underlying stock does not pay any dividends during the life of the options.

Suppose, for instance, we want to know the no-arbitrage stock price given this information:

Call price = \$3.50	Riskless interest rate = 5%
Put price = \$1.00	Time until option expiration = 32 days
Strike price = \$75	

We can rearrange Equation 24.4:

$$\mathrm{S}_0 = C - P + \frac{K}{(1 + r)^T} \tag{24.4a}$$

and plug in the known values:

$$\mathrm{S}_0 = \$3.50 - \$1.00 + \frac{\$75.00}{(1 + .05)^{32/365}} = \$77.18 \tag{24.4b}$$

A simple example will show why the put/call parity relationship must be true. Without arbitrage, equivalent financial claims should sell for the same price. This is called the *law of one price*. Suppose we have the stock and option prices shown in Exhibit 24.6. No matter what the stock price at option expiration, the activities described yield a profit of \$0.31.

Conversely, if the put price is too high relative to the call price, the arbitrageur could write the put, buy the call, sell a share of the stock short, and invest the proceeds from the short sale at the 6 percent interest rate.

The theory of put-call parity indicates that, *when European options are at-the-money and the stock pays no dividends*, relative call prices should exceed relative put prices by an amount approximately equal to the riskless rate of interest for the option term times the stock price.

Consider another situation, this time when the options are not at-the-money. In the previous example, suppose the stock price is \$47 instead of \$50. The \$50 call is out-of-the-money, while the \$50 put now has intrinsic value of \$3. Logically this

Exhibit 24.6 Arbitrage via Option Mispricing

Initial Values:
Stock price (S_0) = \$50
Strike price (K) = \$50
Time until expiration (t) = 6 months
T bill interest rate (r) = 6.00%
Call premium (C) = \$4.75
Put premium (P) = \$3

Theoretical Put Value Given the Call Value:
$P = C - S + K/(1 + R)^T$
$P = \$4.75 - \$50 + \$50(1.06)^{0.5} = \$3.31$
This means the actual call price (\$4.75) is *too high* or that the put price (\$3) is *too low*.

To Exploit the Arbitrage:
Write 1 call @ \$4.75.
Buy 1 put @ \$3.
Buy 1 share @ \$50.
Borrow \$48.56 @ 6.00% for six months.

Stock Price at Option Expiration:

Profit/Loss	**\$0**	**\$50**	**\$100**
From call	4.75	4.75	(45.25)
From put	47.00	(3.00)	(3.00)
From loan	(1.44)	(1.44)	(1.44)
From stock	(50.00)	0.00	50.00
Total	\$0.31	\$0.31	\$0.31

make a difference in the pricing of these two options. Suppose the put actually sells for \$6. A six-month, \$50 call should sell for

$$C = \$6.00 + \$47 - \$50/(1.06)^{0.5} = \$4.44$$

We can confirm this value by constructing another contingency table. If the prices are in equilibrium there should be no arbitrage profit, which Exhibit 24.7 indicates is the case.

Exhibit 24.7 Put/Call Parity Contingency Table

Profit/Loss	**\$0**	**\$50**	**\$100**
From writing call	4.44	4.44	(45.56)
From buying put	44.00	(6.00)	(6.00)
From loan	(1.44)	(1.44)	(1.44)
From buying stock	(47.00)	3.00	53.00
Total	\$0.00	\$0.00	\$0.00

We can learn something else about option pricing from the put-call parity relationship. Suppose we have this information:

Stock price (S) = \$62.13	Riskless interest rate (r) = 6.15%
Strike price (K) = \$60	Time until option expiration (t) = 47 days

We might ask: What is the minimum price for which a \$60 call should sell? Rearranging the put-call parity model, the quantity $C - P$ equals the stock price (S_0) minus the present value of the strike price $K/(1 + r)^t$. Plugging in our values, $C - P = \$62.13 - \$60/(1.0615)^{47/365}$, or \$2.58. The put cannot sell for less than zero, so C must be *at least \$2.58*. Although the put is out-of-the-money, it will still have some time value, so the minimum call premium will actually be \$2.58 plus the time value of the put. Another way of looking at this example is to say that this call must have time value of at least \$2.58 − \$2.13, or about 45 cents.

BINOMIAL OPTION PRICING MODEL[3]

Put-call parity shows that given the put price, we can solve for the call price. We do not need the put value to solve for the call value if we know something about the possible future paths the stock price might take. One way to get some intuition into this fact is via binomial pricing.

We can learn something about this by creating an imaginary capital market to illustrate rational option pricing. Suppose you can invest in U.S. government securities and earn 10 percent over the next year. Stock XYZ currently sells for \$75 per share. For simplicity, assume there are no transaction costs or taxes.

In our hypothetical market, there are only two possible states of the world at year-end: Either the stock will rise to \$100 or it will fall to \$50. There are call options for sale that give you the right to buy the stock a year from today at its current price of \$75. If the stock rises to \$100, the call would be worth \$100 − \$75 = \$25, because it gives its owner the right to buy stock for \$25 less than the market price. If the stock falls to \$50, the option is unattractive and would expire worthless; no one would choose to pay \$75 for stock worth \$50. Exhibit 24.8 illustrates the scenario. For what price should such an option sell?

Returning to pricing our option, one approach to the problem would be to determine the expected value of the stock in one year, from that determine the expected call value, and discount this amount back to a present value. Perhaps an optimistic investor believes there is a 90 percent chance the stock will rise and a 10 percent chance it will fall. With branch probabilities 0.9 and 0.1, the expected stock price is $(0.9 \times \$100) + (0.1 \times \$50) = \$95$. The expected call price is therefore \$95 − \$75 = \$20, with a present value of \$20/1.10, or \$18.18.

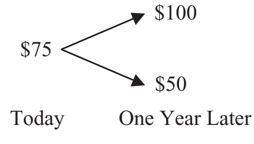


Exhibit 24.8 Possible States of the World

Despite this seemingly logical calculation, such a price presents an arbitrage opportunity that would not prevail for long in a well-functioning marketplace. To see why, consider the two steps an arbitrageur would implement quickly.

1. Buy a share of the stock, spending $75.
2. Write two of these calls at $18.18 each, receiving $36.36. The net investment, then, is $75 − $36.36, or $38.64.

If the stock falls, the options will expire worthless and the portfolio will contain one share of stock worth $50. If the stock rises, the share will be worth $100. The options would be valuable to their owner because each of them permits the purchase of shares worth $100 for only $75. The option holder would exercise this option; the option writer would have to sell two shares for $25 less than their current market value, thereby losing $50. The ending portfolio value would be $100 (the value of the stock) minus the $50 loss on the options, or $50. Exhibit 24.9 shows the possibilities.

This means regardless of whether the stock follows the upper or the lower branch in Exhibit 24.8, the portfolio will be worth $50 in one year. Because the initial cash outlay was $38.64, the $50 terminal value translates into a certain one-year gain of 29.4 percent with no risk. This is inconsistent with a U.S. government risk-free rate of 10 percent; $18.18 cannot be the correct value for the call option.

Suppose a different investor views XYZ's prospects differently and reverses the probabilities for the two branches: She feels the stock has a 10 percent chance of appreciating and a 90 percent chance of falling. Using the prior logic, the stock's expected value is $55 and she concludes the option has no value because the expected future stock price is less than the price the option entitles you to pay. No one would choose to pay $75 for shares worth $55. The arbitrageur, however, offers

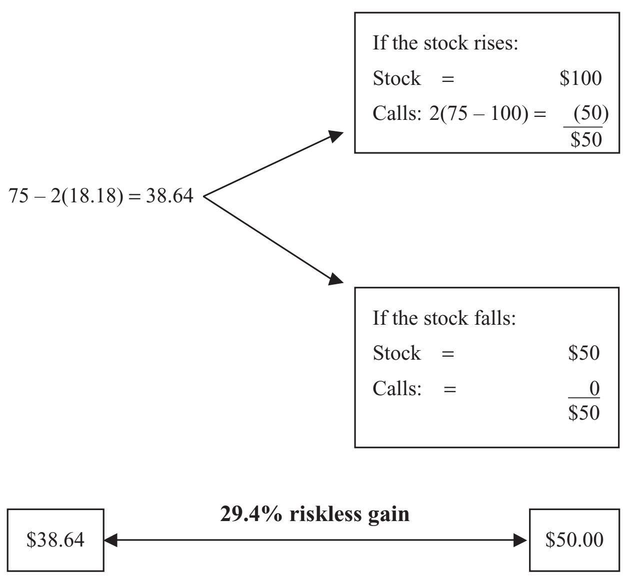


Exhibit 24.9 Portfolio Values

to pay \$1 apiece for these calls if she will write them. Thinking the arbitrageur has erred in his calculations, she agrees to write two such contracts for \$1 each.

Having acquired these two calls, the arbitrageur will then sell one share of the stock short. Selling short involves borrowing a share, selling it, buying a replacement share from someone else at a later time, and replacing the borrowed share. Buying the two calls for \$1 each and selling one share short at \$75 results in a net cash inflow of \$73. As in the prior example, at expiration the portfolio will be worth \$50 regardless of the branch traveled. If the stock goes up, the arbitrageur loses \$25 on the short position when he purchases shares at \$100 to replace those borrowed. However, he profits on each call. The right to buy at \$75 is worth \$25 when the stock price is \$100. Having paid \$1, the gain is \$24 on each call, or \$48 on two of them. The net portfolio gain is 2 × (\$25 − \$1) − \$25 = \$23.

If the stock goes down, the calls expire worthless but the arbitrageur makes \$25 on the short sale for a net \$23 gain, as before. In this case, the return is infinite, as there was no initial cost to this investment; it began with a net inflow. Clearly \$1 is not an equilibrium value for the call either. (See Exhibit 24.10.)

To find out what the call price must be, we can generalize the Exhibit 24.8 example and create an arbitrage portfolio. If the stock rises, the call will be worth \$25; it will be worth \$0 if the stock falls. We can construct a portfolio of stock and options such that the portfolio has the same value regardless of the stock price after one year. One way to do this is to buy one share of stock today and write a quantity of calls we will call N.

If the stock falls, the value of this portfolio will be \$50. The share is worth \$50 and the calls expire worthless. If the stock rises, the portfolio is worth \$100 − \$25N. The stock is worth \$100 and the options are worth \$25 apiece to their owner (or minus \$25 to the person who wrote them). When they are exercised, the option writer will have to sell \$100 stock for \$75 per share, losing \$25 per share. Exhibit 24.11 shows the possibilities.

We can solve for N such that the portfolio value in one year must be \$50. Setting the two possible values equal, \$100 − \$25N = \$50, and N = 2. This means if we buy one share of stock today and write two calls, we know the portfolio will be worth \$50 in one year.[4] In other words, the future value is known and riskless.

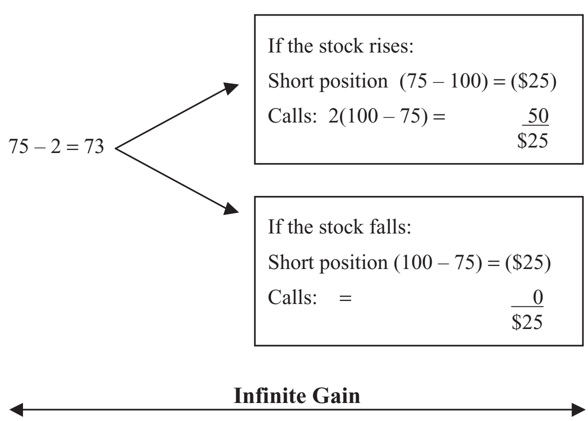


Exhibit 24.10 Portfolio Values

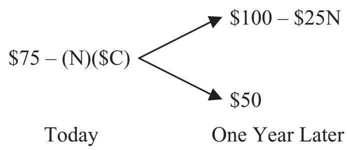


Exhibit 24.11 Portfolio Values

Economic theory requires that an investment with a known future value must earn the riskless rate of interest, which is 10 percent in this example.

Suppose, for instance, the government promises you \$50 in one year (with no risk) and the prevailing one-year riskless rate is 10 percent. This future payment is worth \$50/1.10 = \$45.45 today. If the investment sold for less than this, the return would be greater than 10 percent; if it sold for more, the return would be less than 10 percent. Neither situation is consistent with a riskless rate of 10 percent. The market will price this future cash flow at \$45.45.

Returning to the option example, we know the portfolio will be worth \$50 in one year regardless of the path taken. Given the 10 percent rate, the portfolio must be worth \$45.45 today. Therefore, assuming no arbitrage opportunity exists (\$75 − 2*C*) = \$45.45. Equivalently, (\$75 − 2*C*)(1.10) = \$50.

Solving for *C*, we find the option must sell for \$14.77. This value is independent of the probabilities associated with the two branches. It makes no difference what the probabilities are that the investor assigns to the two branches. At any price above this value, people who buy one share of stock and sell two call options will earn risk-free more than they would by putting money in the bank. Conceptually this means people will be lining up to sell options, and their price would be driven down to \$14.77. Similarly, if the price were lower, there would be risk-free opportunities to option buyers. This discovery is an epiphany for students of derivatives: The price of an option is independent of the expected return on the stock.

This is admittedly a troubling and seemingly illogical result. Modern option pricing theory makes extensive use of continuous time mathematics and the associated calculus. Calculus is about rates of change, especially how one variable changes when another variable changes by a very small amount. You might understand why the future rate of return does not matter by considering a more familiar setting: the speed of an automobile.

Suppose you know that at a precise instant in time a car is traveling at exactly 45 miles per hour. Now ask yourself this: Is the car speeding up or is it slowing down? Upon reflection, you will realize that just knowing the car's speed is an incomplete picture of the situation. The car might be cruising at a constant speed, but it could just as easily have been parked and now be accelerating toward 70. It could also be braking to a stop from highway speed. Regardless of the big picture, at some point the vehicle is traveling exactly 45. Knowing that the car is going 45 miles per hour tells you nothing about its likely future speed. Even the space shuttle, for a very brief instant, travels at 45 mph. The analogy is not perfect, but without going deeply into the mathematics, this may give a hint into why knowing that a call option sells for \$3 tells you nothing about the future path of the stock price. Therefore, the future path of the stock price should not affect the call price.

Note also that the size of the jump and the length of the time period are unimportant. Actual stock prices can change second by second and by very small

Exhibit 24.12 Put-Call Arbitrage

Initial Stock Price = $75		Portfolio Value at Option Expiration	
Activity	**Cash Flow**	**Stock Price = $100**	**Stock Price = $50**
Buy call	−$14.77 (payment)	$25.00	$0.00 (expires worthless)
Write put	+$14.77 (receipt)	$0.00 (expires worthless)	−$25.00
Sell stock short	+$75.00 (receipt)	−$100.00 (cost to close out)	−$50.00 (cost to close out)
Invest $75 in T bills	−$75.00 (payment)	$82.50 (receipt from T bill investment)	$82.50 (receipt from T bill investment)
Totals	$0.00 (net receipt)	$7.50	$7.50

increments. The Black-Scholes model allows infinitesimally small changes in the stock price in continuous time.

PUT PRICING IN THE PRESENCE OF CALL OPTIONS: FURTHER STUDY

Knowing that this call option must sell for $14.77, we can now turn to its counterpart, the put option, and learn something about its equilibrium price. Suppose the put also sells for $14.77. The arbitrageur, knowing that this value is incorrect given the call premium, engages in the transactions shown in Exhibit 24.12.

We see that this series of transactions results in an initial cost of $0.00 and a future portfolio value of $7.50. In other words, we invest no money now but will receive $7.50 in one year. This is the proverbial free lunch.

In fact, the arbitrageur need not even wait a year for this windfall. The arbitrageur could invest only the discounted value of the strike price, or $75/1.10 = $68.18. Exhibit 24.13 shows this would result in an initial cash flow of $6.82 and a portfolio value at option expiration of $0.00.

Exhibit 24.13 shows that with the put and call both selling for $14.77 the arbitrageur could, with no investment and no risk, capture $6.82 today, or, as Exhibit 24.12 shows, capture the future value of this $6.82(1.10) = $7.50 in one year.

Exhibit 24.13 Put-Call Arbitrage

Initial Stock Price = $75		Portfolio Value at Option Expiration	
Activity	**Cash Flow**	**Stock Price = $100**	**Stock Price = $50**
Buy call	−$14.77 (payment)	$25.00	$0.00 (expires worthless)
Write put	+$14.77 (receipt)	$0.00 (expires worthless)	−$25.00
Sell stock short	$75.00 (receipt)	−$100.00 (cost to close out)	−$50.00 (cost to close out)
Invest $68.18 in T bills	−$68.18 (payment)	$75.00 (receipt from T bill investment)	$75.00 (receipt from T-bill investment)
Totals	$6.82 (net receipt)	$0.00	$0.00

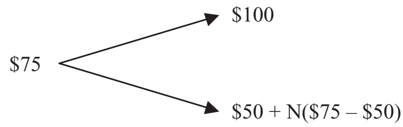


Exhibit 24.14 Binomial Put Pricing

Given a call premium of $14.77, the put cannot also sell for this amount in an arbitrage-free market.

In fact, the put must sell for $\$14.77 - \$6.82 = \$7.95$. This put price would result in an initial cash inflow of $0.00 in Exhibit 24.13. We would expect something that we know to be worthless at all points in the future to also be worthless today. This brings us back to the put/call parity model: $C - P - S + K/(1 + R)^t = 0$

The call premium, put premium, stock price, and strike price form an interrelated securities complex. If you know all variables but one, you can solve for its arbitrage-free value.

BINOMIAL PUT PRICING

The binomial pricing logic used with call options works equally well with put options. Suppose we have the same situation as before: a current stock price of $75 with a one-year-later stock price of either $100 or $50, as in Exhibit 24.14. We can combine at-the-money puts with stock so that the future value of the portfolio is known for certain.

Setting the two portfolio values equal, we find $N = 2$. This means a portfolio composed of one share of stock and two puts will grow risklessly to $100 after one year. This gives us the equation $(\$75 + 2P)(1.10) = \$100$, and $P = \$7.95$. This is exactly the same value we found when we solved for the put premium using put-call parity.

BINOMIAL PRICING WITH ASYMMETRIC BRANCHES

So far the binomial pricing examples have all involved symmetric branches, with the size of the "up" movement equal to the size of the decline. This need not be the case and it makes no difference, because the pricing logic still holds. The only thing we have to do differently is use a quantity of options other than the 2.0 we always have when the branches are symmetric. Exhibit 24.15 shows the possibilities on another option.

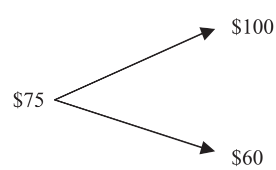


Exhibit 24.15 Portfolio Values

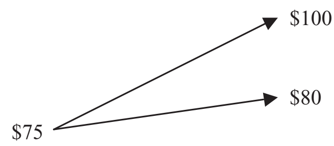


Exhibit 24.16 Binomial Pricing

Here we see that the stock will either rise by \$25 or fall by \$15. Suppose in this example the time period is three months, the annual interest rate is 5.50 percent, and we want to find the value of a \$75 call.

The first step is to determine the number of call options to write in order to create the riskless hedge. Call this number N. If the stock goes up, the portfolio will be worth $\$100 - \$25N$. If it goes down, the portfolio will be worth \$60. Setting these two values equal, we have $\$100 - \$25N = \$60$. Solving, we find $N = 1.6$.

Buying 1 share and writing 1.6 calls, then, will grow risklessly to \$60. Remember that the fractional contract is not a problem here. We could just as easily buy 10 shares and write 16 calls. The valuation equation now becomes $(\$100 - 1.6C)(1.055)^{.25} = \$60$. Solving, the equilibrium call value (C) is \$25.50.

Now let us turn to a slightly more complicated example. This time suppose the branch possibilities are as shown in Exhibit 24.16, with the strike price remaining at \$75, one year until expiration, and the interest rate remaining at 5.50 percent.

If the stock rises to \$100, the portfolio will be worth $\$100 - \$25N$. If the stock rises to \$80, the portfolio value will be $\$80 - \$5N$. Note in the latter case that the options will be in-the-money and, therefore, worth \$5 apiece. Setting the two possible portfolio values equal, $\$100 - \$25N = \$80 - \$5N$. Solving, we find $N = 1.0$, which means the future portfolio value will be \$75.00 regardless of the path. The valuation equation is then $(\$75 - C)(1 + .055) = \$75.00$, and $C = \$3.91$.

Calls are worth less the higher their strike price. What would a \$78 call on the stock in Exhibit 24.16 be worth? Solving for N, we set $\$100 - \$22N = \$80 - 2N$ and find $N = 1.0$. The valuation equation is $(\$75 - C)(1.055) = \$78$, and $C = \$1.07$. This is less than the premium for the \$75 call, as expected.

EFFECT OF TIME

With options, more time until expiration means more value. Suppose the \$78 call in the last example had 18 months until expiration instead of 12. All that changes in the valuation equation is the time period:

$$(\$75 - C)(1.055)^{1.5} = \$78$$

Solving, $C = \$3.02$.

Now suppose the time until expiration is six months; we would expect the option premium to decline.

$$(\$75 - C)(1.055)^{0.5} = \$78$$

Solving, C = minus \$0.94.

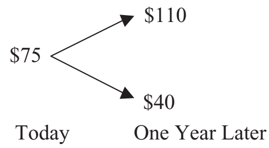


Exhibit 24.17 Possible States of the World

We know that an option cannot sell for less than zero, so something is wrong here. The problem is that we have introduced arbitrage into our market. We assumed the riskless interest rate is 5.50 percent per year and then constructed a portfolio with a known future value of $78 and a starting stock price of $75. If the call initially sold for zero, an increase in portfolio value from $75 to $78 in six months is an annual interest rate of 8.16 percent with no risk. If the call sold for more than zero, the rate of return would be even higher. This is inconsistent with a riskless rate of 5.50 percent, as we assumed.

This example illustrates the fact that you cannot just assume any conditions you want. If your assumed market contains arbitrage, option pricing principles will ferret it out.

EFFECT OF VOLATILITY

Perhaps in our hypothetical market there is another firm whose stock also sells for $75 per share, but this stock is more volatile: In the next year it will either rise to $110 or fall to $40, as in Exhibit 24.17.

Using the same pricing procedure as in the earlier examples, a one-year, $75 strike price call option on this stock should sell for $19.32. This is the solution to the equation $(75 - 2C)(1.10) = 40$. Using the put-call parity model, the equilibrium value of the corresponding put is $C - S + K/(1 + r)^t$, or $\$19.32 - 75 + 68.18 = \$12.50$. The higher premium associated with the greater volatility shows why options on technology stocks sell for a "higher price" than options on similarly priced retail food stores or electric utilities.

You can also get some idea as to why options on volatile assets are expensive by thinking about the pricing of an insurance policy. Drivers with a history of accidents and speeding tickets pay a higher premium than someone with a perfect driving record. Similarly, an insurance policy protecting you against a very remote possibility (such as flight insurance) is inexpensive. A technology stock that often moves up or down 5 percent per day is susceptible to an "accident" and will have expensive options, while an electric utility is likely to be much more stable (and have lower option premiums).

INTUITION INTO BLACK-SCHOLES

The fact that future security prices are not limited to only two values in no way attenuates the usefulness of the binomial pricing model. We can simultaneously make the size of the jumps very small and the time interval very short. There are theoretically an infinite number of future states of the world. By making the jumps infinitely small and the time span infinitely short, we move into the world

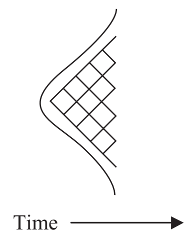


Exhibit 24.18 Moving from the Binomial to Continuous Time

of continuous time calculus and the Black-Scholes model for which Myron Scholes received the Nobel Prize, as Fisher Black certainly would have, had he lived. Exhibit 24.18 shows how the one-period tree diagrams can be extended to multiple periods and how you might visualize a normal distribution about the range of future states of the world.

The arguments in the Black-Scholes model are the current stock price, the option strike price, the remaining time until option expiration, the interest rate, and the anticipated volatility of the underlying asset. The pricing logic remains: A riskless investment should earn the riskless rate of interest. If this is not the case, arbitrageurs will quickly transact so as to move prices to their equilibrium relationship. As the earlier example showed, it is not necessary to know the expected rate of return on the stock in order to find the equilibrium value of the right to buy the stock. This is just as true in continuous as in discrete time.

ENDNOTES

1. The vertical axis shows profit or loss, while the horizontal axis shows the stock price at option expiration. The bend in the diagram occurs at the option strike price.
2. An "investment" requires a cash outlay. A simultaneous long position in stock and short position in the same stock involves no outlay of funds, is therefore not an investment, and does not earn the riskless rate of interest.
3. Much of the material in this section comes from Strong and Buonocristiani (2000).
4. Note that this is not a prescribed investment strategy. No one would choose to engage in a strategy that is guaranteed to lose money. We are merely solving for a value that must exist in the absence of arbitrage.

REFERENCES

Stoll, H. 1969. "The relationship between put and call option prices," *Journal of Finance* 24: 5 (December): 801–824.

Strong, R. A., and J. Buonocristiani. 2000. "Wall Street Profits, Arbitrage, and the Pricing of Stock Options," *Chance* (Summer): 20–24.

FURTHER READING

Black, F., and M. Scholes. 1973. "The Valuation of Options and Corporate Liabilities," *Journal of Political Economy* 81, no. 3 (May): 637–654.

Black, F., and M. Scholes. 1989. "How We Came Up with the Option Formula," *Journal of Portfolio Management* 15, no. 2 (Winter): 4–8.

Grant, D., G. Vora, and D. Weeks. 1995. “Teaching Option Valuation: from Simple Discrete Distributions to Black/Scholes via Monte Carlo Simulation,” *Financial Practice and Education* 5, no. 2 (Fall/Winter): 149–155.

Klemkosky, R., and B. Resnick. 1980. “An Ex Ante Analysis of Put–Call Parity,” *Journal of Financial Economics* 8, no. 4 (December): 363–378.

Rendleman, R., and B. Bartter. 1979. “Two-State Option Pricing,” *Journal of Finance* 34, no. 5 (December): 1093–1110.

Rubenstein, M., and H. Leland. 1981. “Replicating Options with Positions in Stock and Cash,” *Financial Analysts Journal* 37, no. 4 (July/August): 63–72.

ABOUT THE AUTHOR

Bob Strong is Foundation Professor of Investment Education and professor of finance at the University of Maine. The UMaine Alumni Association selected him as the 2005 Distinguished Maine Professor. The Carnegie Foundation named him the 2007 Maine Professor of the Year.

He earned a BS in engineering from the United States Military Academy, an MSBA from Boston University, and a PhD in finance from Penn State. He was deputy director of the Summer Economics Program at Harvard University from 1997 to 1999. He is a Chartered Financial Analyst and has been a conference speaker for the Chicago Board Options Exchange, the Chicago Board of Trade, and the American Stock Exchange.

His three textbooks—*Investments*, *Portfolio Management*, and *Derivatives*—are used at over 100 universities. Professor Strong is past president of the Northeast Business and Economics Association and the Maine CFA Society and is an honorary captain in the Maine State Police.

CHAPTER 25

The Pricing of Forward and Futures Contracts

DAVID DUBOFSKY
Professor of Finance and Associate Dean for Research, College of Business, University of Louisville

There are spot prices and there are forward/futures prices. The spot price of something is the price you pay today for taking delivery today. The forward price or futures price is an agreed-upon price that will be paid on a future delivery day. On the delivery day, the party who is long a forward contract will pay the originally agreed-upon price to the party who is short that forward contract. In exchange, the party who is short the forward contract will deliver the underlying asset to the party who is long the forward contract. Obviously, spot prices and forward/futures prices can be different.

Arbitrage is one of the strongest pricing concepts in financial theory. Simply put, prices should not permit arbitrage, which is a set of trades that requires no initial cost or cash outflow but also leads to a cash inflow (a profit) in the future. Some futures and forward prices are determined via the force of arbitrage. Futures and forwards on other underlying assets are determined using models that are less precise; that is, there is more latitude about what the futures price should be because it is affected by unobservable and/or immeasurable factors.

There are two basic models for estimating a theoretically correct futures price: the cost of carry model and the expectations model. The cost of carry model is used to price futures contracts on financial assets (stocks, currencies, interest rates and bonds, and gold). It can also be adapted (with a concept known as the convenience yield) to price futures contracts on commodities. With some underlying assets, the cost of carry model is enforced by arbitrage. The expectations model is used to estimate futures prices for commodities and nonstorable underlying assets, such as electricity.

For ease of exposition, we use the terms *futures contract* and *futures price* rather than *forward contract* and *forward price*. Later in the chapter, we address the issue of whether a futures price and a forward price should be the same.

To make our lives easier, we ignore transactions costs (commissions, bid-ask spreads, borrowing and lending interest rate differentials, and price pressure/impact from a trade). In other words, we assume that you can buy and sell the underlying asset at the same price, go long and short the futures contract at the same price, and borrow and lend at the same interest rate. We assume you can sell the underlying asset short and get full use of the proceeds. We ignore taxes. We ignore any possibility that you or your counterparty will fail to abide by the terms of your contract. (Hence we ignore any practice that reduces default risk such as margin, collateral, or marking to market.) Finally, we assume that markets operate sufficiently well that there are no arbitrage opportunities. All rational wealth-maximizing traders will trade if the net initial cash flow is zero, and they know that they will get paid (receive money) in the future.

COST OF CARRY MODEL

Cash-and-carry arbitrage is the foundation of the cost of carry pricing model when the underlying asset is a financial asset. The set of trades that make up cash-and-carry arbitrage is

- Borrow
- Buy the underlying asset
- Sell (go short) a futures contract

This set of trades has the arbitrageur borrowing in one market and lending in another market (the market of the underlying asset). The act of buying the underlying asset is a loan (a cash outflow). The loan will be repaid with interest on the delivery day of the futures contract. Selling the futures contract locks in the selling price; it locks in the repayment on the loan. Arbitrage exists if you can borrow in one market at an interest rate that is below the riskless lending rate that exists in another market.

Cash-and-carry arbitrage establishes a maximum futures price. If the futures price is too high, then this set of trades will lead to an arbitrage profit.

Reverse cash-and-carry arbitrage sets a minimum futures price. If the futures price is too low, then the following set of trades will lead to an arbitrage profit:

- Sell the underlying asset (either sell it short or sell it out of your inventory).
- Lend the proceeds from the sale.
- Go long a futures contract.

We denote S as the spot price of the underlying asset, and F as the futures price. The subscript 0 denotes "today" and T denotes the delivery day. The interest rate is r per year, and the initial time until delivery is T years. We assume continuous compounding so that the future value of a dollar that will be received at time T is e^{rT}.[1]

Exhibit 25.1a illustrates cash-and-carry arbitrage, and Exhibit 25.1b illustrates reverse cash-and-carry arbitrage. In each panel, the trades are stated in the first column. The second column is an algebraic illustration of the cash flows (where a + represents a cash inflow and a – denotes a cash outflow). The third column provides a numerical example for the initial cash flows (today). The last two columns provide two (out of an infinite number of) possible outcomes: one in which the futures price rises and the other in which the futures price declines. Either way, the same cash flow results on the delivery day.

Exhibit 25.1a Cash-and Carry Arbitrage

		$S = 50$ $F_0 = 52$ $r = 8\%$ $T = 3$ months	
Today			
Borrow	$+S$	$+50$	
Buy the underlying asset	$-S$	-50	
Go short futures contract with a futures price of F_0	0	0	
Total initial cash flow	0	0	
AT DELIVERY			
Sell the underlying asset	$+S_T(+F_T)$	$S_T = F_T = 54$ $+54$	$S_T = F_T = 46$ $+46$
Profit or loss on the futures contract	$F_0 - F_T$	$52 - 54$	$52 - 46$
Repay loan	$-Se^{rT}$	$-50e^{(0.08)(0.25)}$ $= -51.01$	$-50e^{(0.08)(0.25)}$ $= -51.01$
Total cash flow at delivery	$F_0 - Se^{rT}$	$+0.99$	$+0.99$

Exhibit 25.1b

		$S = 50$ $F_0 = 50.5$ $r = 8\%$ $T = 3$ months	
TODAY			
Sell the underlying asset	$+S$	$+50$	
Lend the proceeds from the sale	$-S$	-50	
go long futures contract with a futures price of F_0	0	0	
Total initial cash flow	0	0	
AT DELIVERY			
Buy the underlying asset	$-S_T(= -F_T)$	$S_T = F_T = 54$ -54	$S_T = F_T = 46$ -46
Profit or loss on the futures contract	$+F_T - F_0$	$54 - 50.5$	$46 - 50.5$
Receive loan principal and interest	$+Se^{rT}$	$+50e^{(0.08)(0.25)}$ $= +51.01$	$+50e^{(0.08)(0.25)}$ $= +51.01$
Total cash flow at delivery	$-F_0 + Se^{rT}$	$+0.51$	$+0.51$

In Exhibit 25.1a, the initial futures price is too high. The cash-and-carry arbitrage trades have you borrow to buy the spot underlying asset and sell the overpriced futures contract. There is no initial cash flow. No matter what happens at delivery, the arbitrageur realizes a profit of 0.99.

In Exhibit 25.1b, the initial futures price is too low. Reverse cash-and-carry arbitrage requires that you sell the underlying asset, lend the proceeds, and buy the cheap futures contract. Again, there is no initial cash flow. The arbitrageur receives a cash inflow of 0.51 regardless of the delivery day futures price.

Because of convergence, $F_T = S_T$; the futures price must equal the spot price on the delivery day.

We conclude that in order for there to be no arbitrage opportunity, the futures price must equal:[2]

$$F = Se^{rT}$$

CARRY RETURN

The previous example and discussion of cash-and-carry arbitrage applied to underlying financial assets that provide no carry return. Examples of such assets are gold and non–dividend-paying stocks.

A carry return is a monetary benefit from actually owning the underlying asset. All else being equal, a carry return will lower the futures price relative to the spot price. (In contrast, a carry cost increases the futures price relative to the spot price.) Specific examples of a carry return are:

- *Dividends,* for dividend-paying stocks. More accurately, it is the future value of the dividends received between today and the futures contract's delivery day.
- *Coupon payments,* for a bond (plus the interest that can be earned on those coupons).
- *Interest,* denominated in the foreign currency, for a unit of foreign currency. More accurately, it is the future value of the interest that can be earned on the foreign currency, when converted back to the home currency (dollars, for U.S.-based traders).

For stock index futures, or a futures contract on a dividend-yielding stock, define d as the annualized dividend yield (annual dividends divided by price), and the theoretical futures price is:[3]

$$F = Se^{(r-d)T}$$

This model works best when dividends are paid smoothly by the stocks in the index, throughout the life of the futures contract. It also assumes that we are reasonably certain that all future dividends will be paid.

For example, assume that the spot Standard & Poor's (S&P) 500 index is at 1300. The annual dividend yield on the S&P 500 is 2 percent, and the six-month interest rate is 3.5 percent per annum. The theoretical S&P 500 futures price for delivery six months hence is

$$F = 1300e^{(0.035-0.02)0.5} = 1300e^{0.0075} = 1300(1.007528) = 1309.79$$

In discrete terms, the theoretical futures pricing model for a dividend paying stock or stock index is

$$F = S(1+r)^T - FV(\text{divs})$$

where $FV(\text{divs})$future value of the cash dividends paid between now and delivery

Equivalently, this discrete model is:

$$F = [S - PV(\text{divs})](1+r)^T$$

where $PV(\text{divs})$present value of the dividends paid prior to delivery

Consider this example that illustrates the nature of discrete "lumpy" dividends. A futures contract exists on an individual stock with a stock price of 50. The delivery day is two months hence. The riskless interest rate is 3 percent per annum. For simplicity, assume the stock will trade ex-dividend and pay its dividend on the same day, one month from today. The dividend amount is \$0.80 per share.

Using the discrete pricing model, the theoretical futures price is 49.445:

$$F = [S - PV(\text{divs})](1+r)^T = [50 - 0.80/(1.03)^{0.0833}](1.03)^{0.1667} = 49.445$$

For a foreign currency, define f as the foreign interest rate and r as the domestic interest rate. The theoretical futures price is:[4]

$$F = Se^{(r-f)T}$$

For example, suppose that the spot price of a euro is \$1.42/€. Suppose that the annual interest rate in the United States is 5 percent, and the annual interest rate for borrowing and lending euros is 3 percent. The theoretical futures price for a contract to deliver euros six months hence is $F = 1.42e^{(0.05-0.03)0.5} = 1.42e^{0.01} = (1.42)(1.01005) = \$1.4343/€$.

Another carry return that is sometimes relevant in the futures pricing of some commodities is called the *lease rate*. If the owner of the underlying asset knows that she will not need to use it in the near future (prior to the futures contract's delivery date), she can lend it to someone and be repaid "interest" in the form of additional product. For example, the owner of 100 ounces of gold can lease it to another party for a year and get repaid 100.5 ounces of gold a year later. The lease rate is a carry return, and if expressed in percentage terms, it should be treated just like a dividend, d.

COMMODITY FUTURES

For investment assets, the only relevant carry cost is the interest forgone by purchasing (a cash outflow) the underlying asset. For commodities, other carry costs are relevant, such as the costs of storing and insuring the commodity. There may be risk of spoilage or theft. Like financial carrying costs (interest), the cost of physically storing the underlying asset increases the futures price, all else being equal. Define c as the (future value of the) percentage of the spot price that must be paid to store the underlying asset until delivery, and the futures pricing model is

$$F = Se^{(r+c)T}$$

Alternatively, if C is the present value of all of the physical storage costs, then

$$F = (S + C)e^{rT}$$

CONVENIENCE YIELD

Recall that reverse cash-and-carry arbitrage requires either that the underlying asset be sold out of inventory or be sold short. Reverse cash-and-carry arbitrage establishes the lower bound for the futures price; futures prices cannot be so low that arbitrageurs will sell the underlying asset, lend the proceeds, and buy the cheap futures contract.

Often, however, all owners of the underlying asset need it as part of their businesses. Oil refiners own crude oil, but they will not sell it to engage in reverse cash-and-carry arbitrage even when the opportunity to realize a riskless arbitrage profit exists. They will not lend their crude oil to a short seller either. A cereal manufacturer will not sell short the raw material (grains) it needs to produce its final product.

The convenience yield is the concept that reconciles this reluctance or inability to sell (short) with the reality of futures pricing. The convenience yield is the unobservable variable that measures the marginal benefit of owning the underlying asset.[5] The convenience yield lowers the futures price that can exist, because reverse cash-and-carry arbitrage is not taking place.

If y is defined as the convenience yield, as a percentage of the price of the underlying asset, then the lower bound for no-arbitrage futures pricing becomes

$$F = Se^{(r-y)T}$$

If storage costs and the convenience yield are both relevant, the model for estimating the theoretical commodity futures price is

$$F = (S + C)e^{(r-y)T}$$

We must stress that y is not observable. You cannot find a table that displays the value of y in your newspaper. But we know that owners of many underlying assets never sell their inventories, even when futures prices are much below spot prices and when futures prices for more distant delivery are lower than futures prices for more immediate delivery.[6] Thus, the convenience yield does exist.

Kaldor (1939) and Working (1948, 1949) originally argued that the convenience yield should increase as spot inventories decrease. While Brennan's (1958) early empirical research failed to support this inventory hypothesis, subsequent research has been more supportive. Pindyck (2001) argues that the convenience yield depends on the demand for storage (which is often seasonal), current supply and demand conditions in the spot market, and the volatility of the spot price. He finds support for these factors, and also finds that the convenience yield tends to be high when the spot price is unusually high.

DELIVERY OPTIONS

Some futures contracts, such as corn, soybeans, wheat, crude oil, and Treasury bonds and notes, convey delivery options to the seller. These delivery options might give the seller:

- A range of delivery dates to make delivery (a "timing option").
- A choice of exactly what type/grade/quality of the underlying asset that will be delivered (a "quality option").
- The ability to decide to make delivery at time t, and receive a payment based on the closing futures price that existed hours or even days prior to time t. (These have been termed the *wild card option* and the *end-of-the-month option*, respectively.)
- A range of delivery locations where the underlying asset will be delivered (a "location option").

A party who is short a corn futures contract that trades on the Chicago Board of Trade can declare his intent to deliver on any day during the delivery month, can deliver different grades of corn at specified price discounts or premia (relative to the actual futures price), and can deliver to any one of three alternative locations (again, with prespecified price discounts or premia).

Exhibit 25.2 illustrates the pricing differentials for different grades of oats and for different delivery locations for wheat.

Exhibit 25.2 Grade Differentials for Oats, and Location Differentials for Wheat

Oats Grade Differentials	
No. 1 Extra Heavy Oats	At 7 cents per bushel over contract price.
No. 2 Extra Heavy Oats	At 4 cents per bushel over contract price.
No. 1 Heavy Oats	At 3 cents per bushel over contract price.
No. 2 Heavy Oats	At contract price.
No. 1 Oats	At contract price.
No. 2 Oats (36 Ib. minimum test weight)	At 3 cents per bushel under contract price.
No. 2 Oats (34 Ib. minimum test weight)	At 6 cents per bushel under contract price.

Wheat Location Differentials

In accordance with the provisions of Rule 1041.00C, wheat in regular warehouses located within the Chicago Switching District, the Burns Harbor, Indiana, Switching District or the Toledo, Ohio, Switching District may be delivered in satisfaction of wheat futures contracts at contract price, subject to the differentials for class and grade outlined above. Only No. 1 Soft Red Winter and No. 2 Soft Red Winter Wheat in regular warehouses located within the St. Louis-East St. Louis and Alton Switching districts may be delivered in satisfaction of wheat futures contracts at a premium of 10 cents per bushel over contract price, subject to the differentials for class and grade outlined above.

All of these delivery options belong to the individual who is short futures. They are valuable options, because the short can choose the method of delivering that costs the least for him or her. Because of this, all else being equal, it is better to be short futures than long futures when you own these options as the short. Hence, actual futures prices may be *slightly* below the prices predicted by any of the above models. The "discount" equals the value of the delivery options owned by the party who is short the futures contract.

Considerable research has gone into estimating these different delivery options.[7]

INTEREST RATE FUTURES AND FORWARDS: EURODOLLAR FUTURES AND FORWARD RATE AGREEMENTS

The cost of carry model determines the prices of futures and forward contracts on interest rates and fixed income securities. But some peculiarities necessitate that these contracts be discussed separately.

First we discuss how to estimate the prices of forward rate agreements (FRAs), which are forward contracts on interest rates, and Eurodollar futures contracts, which trade on the Chicago Mercantile Exchange (CME). In the next section, we will discuss the pricing of Treasury Bond and Treasury Note futures contract, which trade on the Chicago Board of Trade (CBOT).

We define time 0 as today, $t1$ as the time until delivery, and $t2$ as the time until the underlying asset matures. Let us clarify this using FRA notation. A $t1$ X $t2$ FRA is an agreement to borrow (if you buy the FRA) from time $t1$ until time $t2$; the seller of a FRA agrees to lend from time $t1$ until time $t2$. When discussing FRAs, $t1$ and $t2$ are usually in months. Thus, if today is March 1, 2009, a 2 X 6 FRA locks in an interest rate from May 1 until August 31, 2009, a period of four months ($t2 - t1$) beginning two months hence. The shorter-term interest rate is denoted $r(t1)$ and the longer-term interest rate is denoted $r(t2)$. The forward interest rate is $fr(t1, t2)$.

CME Eurodollar futures contracts are basically $t1$ X $t1$+3 FRAs, where $t1$ is in months. They serve to lock in interest rates for forward borrowing and lending over a future three-month (90 days) period, beginning at time $t1$. Time $t1$, the delivery day, is the second business day in London prior to the third Wednesday of the delivery month. Eurodollar futures prices are quoted as

$$100 - \text{3-month forward LIBOR}$$

Thus, if the futures price is 95.48, then the forward LIBOR ($fr(t1, t1+3)$) is 4.52 percent. If on September 21, 2008, you go long December Eurodollar futures, you have effectively agreed to buy \$1 million of 3-month Eurodollar time deposits and will take delivery of those securities on December 15, 2008. We say "effectively" because Eurodollar futures contracts are cash settled; there is no delivery of the underlying asset.

Forward interest rates, and hence the forward/futures prices of FRAs and Eurodollar futures contracts, are determined via simple cash and carry arbitrage.

You borrow (until delivery at time $t1$), buy the underlying asset (a zero-coupon debt security that matures at time $t2$), and sell the forward/futures contract.

For example, consider a 2 X 6 FRA. To estimate the 4-month forward interest rate that begins two months hence, and ends six months hence ($fr(t1, t2) = fr(2, 6)$), you would do this today:

1. Borrow for two months at the two-month interest rate, $r(t1) = r(2)$.
2. Lend for six months at the six-month interest rate, $r(t2) = r(6)$.

When you buy the six-month security, you are lending to the issuer for six months.

We will not go through the mathematics demonstrating how borrowing and lending at the two interest rates is equivalent to borrowing to buy the underlying asset. We instead offer two formulas that compute forward rate. If time $t2$ is longer than a year, it is usually best to use this formula:

$$(1 + r(t2))^{t2} = (1 + r(t1))^{t1}(1 + fr(t1, t2))^{t2-t1}$$

where $t1$ and $t2$ are in years. If time $t2$ is less than a year, it is best to use unannualized interest rates and drop the exponents. Notationally, we will use h as an unannualized interest rate (think of it as a holding period interest rate); that is, $h = r\,X(t/365)$ when t is in days, or $h = r\,X(t/12)$ when t is in months. For example, if $r = 5$ percent per annum, and we want to know the unannualized interest rate over a 3-month period, then $h = 0.05$ X $(3/12) = 0.0125 = 1.25\%$. The formula to compute a forward interest rate is then:

$$(1 + h(t2)) = (1 + h(t1))(1 + fh(t1, t2))$$

After you have determined $fh(t1, t2)$, you must annualize it to get the annualized forward rate.

For example, suppose you want to determine the price of a 24 X 33 FRA, which covers forward borrowing or lending over a 9-month period beginning 24 months hence and ending 33 months hence. Suppose the 2-year interest rate is 5.2 percent and the interest rate for securities maturing 2.75 years hence is 5.4 percent. Then

$$\begin{aligned}
&(1.054)^{2.75} = (1.052)^2(1 + fr(2, 2.75)^{0.75}\\
&1.155611 = 1.106704(1 + fr(2, 2.75)^{0.75}\\
&(1 + fr(2, 2.75)^{0.75} = 1.044192\\
&fr(2, 2.75) = 5.935\%
\end{aligned}$$

The price of the 24 X 33 FRA is 5.935 percent.

As another example, suppose it is October 1, 2008. The delivery day for the December 2008 Eurodollar futures contract is December 15, 2008, 75 days later; that is, $t1 = 75$ days. Because the Eurodollar futures contract is a 90-day contract, $t2 = 165$ days. Assume that the 75-day interest rate is 5 percent per annum, so $h(t1) = h(75) = 0.05 \times 75/365 = 0.010274$. Assume that the 165-day interest rate is 4.8 percent per annum, so $h(t2) = 0.048 \times 165/365 = 0.0216986$. The steps to calculate the forward rate from $t1 = 75$ to $t2 = 165$ days, $fh(75, 165)$ are:

$$
\begin{aligned}
&(1 + h(t2)) = (1 + h(t1))(1 + fh(t1, t2))\\
&1.0216986 = 1.010274(1 + fh(t1, t2))\\
&1.0111265 = 1 + fh(t1, t2)\\
&fh(t1, t2) = fh(75, 165) = 0.0111265 = 1.11265\%
\end{aligned}
$$

This unannualized rate of 1.11265% is annualized as follows:[8]

$$1.11265\% \times 360/90 = 4.4506\%$$

And the theoretical futures price for the Eurodollar futures contract is 100 – 4.45 = 95.55.

INTEREST RATE FUTURES AND FORWARDS: TREASURY BOND AND TREASURY NOTE FUTURES

Estimating the theoretically correct Treasury (T) bond and T note futures price may be more complex than just about any other futures contract. Conceptually, the cost of carry model determines the prices of T bond and T note futures contract. You borrow until the delivery date, buy the underlying asset (a deliverable Treasury security), and sell futures in order to perform cash and carry arbitrage.

However the CBOT T bond and T note futures contracts are very complicated. One complication is that they permit any one of several eligible Treasury securities to be delivered (a quality option). Here is what can be delivered by the party who is short a T bond futures contract:

> *U.S. Treasury bonds that, if callable, are not callable for at least 15 years from the first day of the delivery month or, if not callable, have a maturity of at least 15 years from the first day of the delivery month. The invoice price equals the futures settlement price times a conversion factor plus accrued interest. The conversion factor is the price of the delivered bond ($1 par value) to yield 6 percent.*[9]

The complete list of Treasuries that satisfy the specifications to be "deliverable" can be found at the CBOT Web site: www.cbot.com/cbot/pub/page/0,3181,830,00.html.

The CBOT adjusts the invoice amount (the amount that the short will be paid), depending on exactly which security is selected for delivery by the short. The adjustment is made using a conversion factor. Conversion factors are also available at the above CBOT Web site. The higher the conversion factor, the more the short will be paid for that particular bond, when it is delivered in a particular delivery month. Typically, the more valuable the bond, the greater the conversion factor. In this way, if the short chooses a cheap (low price or low value) bond to deliver, she will be paid less. If she chooses to deliver an expensive (high price or high value) bond, she will be paid more.

So, the party who is short a T bond or a T note futures contract has a valuable quality option. She can select any one of the deliverable Treasuries. The markets know that the short will most likely choose the eligible security that is cheaper to deliver than the others. The cheapest-to-deliver security is the one that costs the least in the cash market but that also gets the deliverer (the party who is short the contract) the highest price upon delivery. T bond and T note futures contract prices tend to track the cheapest to deliver treasury security.[10]

For simplicity, let us assume that the cheapest-to-deliver bond will pay no coupon interest between today and the delivery day. When the underlying asset is a Treasury providing a carry return, a version of the theoretical futures pricing model derived via cash-and-carry arbitrage is:

$$F^* = Se^{rT} - CR$$

where F^* = futures price that is adjusted for the conversion factor. That is, $F^* = (cv)(F)$, where F is the quoted futures price that you will see in the newspaper or at www.cbot.com, and cv is the conversion factor for the cheapest-to-deliver Treasury.

S = spot price of the cheapest-to-deliver Treasury security (quoted without accrued interest).

r = riskless annual interest rate for the period between today and the delivery date.

T = time to delivery in years.

CR = t carry return, which equals the difference between the accrued interest received upon delivery and the initial accrued interest paid when the treasury was purchased. That is, $CR = AI_T - AI_0$.[11]

Thus, the theoretical futures price for a T bond or T note futures contract is:

$$F = \frac{Se^{rT} - (AI_T - AI_0)}{cv}$$

This model ignores the effects of two other options that the short party owns: the end-of-day ("wild card") option and the end-of-the-month option. The wild card option gives the short the right to declare his intent to deliver as late as 8 PM on any day of the delivery month. That day's futures settlement price will then be part of the ultimate invoice amount. The short then has until 2 PM on the following day to decide which Treasury to deliver. This determines the invoice amount. The short can exploit changing market conditions between the time he declared his intention to deliver and 2 PM the next day. Kane and Marcus (1986b) value the wild card option.

The end-of-the-month option exists because the last trading day of T bond futures and T note futures contracts is the seventh business day prior to the last business day of the delivery month. The invoice amount is determined by the settlement futures price on the last day of trading. The short has the last seven days to decide when she wants to deliver and which Treasury to deliver.

SHOULD FUTURES AND FORWARD PRICES BE THE SAME?

In all of our prior discussion, we have used the term *futures price* (not *forward price*). A natural question to ask is whether all of our models hold for forward pricing as well. There are some differences between futures contracts and forward contracts, and one of them in particular could result in the two prices being different: Recall that futures contracts are marked to market daily, while forward contracts typically are settled up only on the delivery date.

Consider going long a *forward* contract to buy gold one year hence. The forward price that you agree to is $920/ounce. If, at delivery, the spot price (and hence the forward price, because of *convergence*) is $950/ounce, then on the delivery date, your profit of $30/ounce is realized. You pay $920 and acquire something worth $950.

In contrast, with a futures contract, the futures price changes each day, and hence each day you realize a relatively small cash inflow (if the futures price rises and you have a long futures position), or a relative small cash outflow (if you are long and the futures price declines). If you add up each of these daily cash resettlement cash flows, you will find that your net cash inflow is $30; this is the sum of the daily mark-to-market cash flows. The difference between settlement of futures and forwards is that forwards are settled at delivery while futures are settled up every day.

While the sum of the daily resettlement cash flows will equal the forward contract profit of $30, the net $30 profit on the long futures position ignores the time value of money. That is, you may have had to borrow to pay the cash outflows when the futures price fell or had the opportunity to invest the cash inflows when the futures price rose.

Academic research has concluded that this timing difference in the cash flows can create a theoretical difference between futures prices and forward prices. To be clear, we are talking about the difference between the price of a futures contract and a forward contract on the same underlying asset with the same delivery date.

Specifically, if interest rates tend to be higher on days that futures prices rise, then the party who is long the futures contract will be able to lend those mark-to-market cash inflows at higher rates; if interest rates tend to be lower on days that futures prices fall, then the party who is long will be able to borrow at a lower interest rate to make those required daily resettlement cash payments. Thus, long futures positions benefit when changes in futures prices and changes in interest rates are positively correlated. If $corr(\Delta F, \Delta r) > 0$, we would expect futures prices to be somewhat greater than forward prices. A trader will prefer to go long futures rather than go long a forward contract, when he believes that interest rates will tend to rise when futures prices rise and that interest rates will tend to fall when futures prices fall. This preference for long futures positions when $corr(\Delta F, \Delta r) > 0$ will result in futures prices slightly exceeding forward prices.

However, a trader will prefer to go short futures rather than go short forward contracts when changes in futures prices and changes in interest rates are negatively correlated. Thus, if $corr(\Delta F, \Delta r) < 0$, we would expect futures prices to be somewhat lower than forward prices.

This theory behind futures and forward price differentials was first advanced by Cox, Ingersoll and Ross (1981) and by Jarrow and Oldfield (1981).

Other factors that may create small differences between futures prices and forward prices include taxes, transactions costs, liquidity, the risk that the counterparty to the contract will default, and the costs of reducing counterparty default risk exposure (margin, marking to market, insurance, etc.). But in general, the total impact of these factors typically will be small, so that generally, we can safely assume that futures prices and forward prices are the same.

Some empirical work has been done to test whether forward and futures prices are the same and whether the correlation of futures price changes and interest rate changes can explain any price differences that do exist. It is difficult to test this relationship. Both the futures and the forward contract must have the same delivery date, delivery location, and underlying asset. Prices must be available for both contracts at the same point in time.

Grinblatt and Jegadeesh (1996) examine Eurodollar futures and forwards and conclude that since 1987, LIBOR futures and forward rates are virtually the same. Other types of underlying assets have also been investigated to determine if forward and futures prices differ: copper and silver (French, 1983), foreign exchange (Chang and Chang, 1990; Cornell and Reinganum, 1981), and stock indices (Cornell and French, 1983; MacKinlay and Ramaswamy, 1988). Park and Chen (1985) examine several commodities, Treasury bonds, and currencies.

EXPECTATIONS MODEL: AN ALTERNATIVE THEORY FOR THE PRICING OF FORWARDS AND FUTURES

Fama and French (1987) begin their paper by presenting two alternative theories that explain futures prices. They call the first the theory of storage, which is the cost of carry model that we have already covered. They attribute this theory to early works by Kaldor (1939), Working (1948, 1949), Brennan (1958) and Telser (1958).

The second model states that the futures price equals the expected delivery-day spot price plus/minus a risk premium, and they attribute its genesis to early research by Cootner (1960), Dusak (1973), and others. Before we start, note that Fama and French (1987) cite a dozen papers and conclude that there is little agreement whether futures can predict delivery day spot prices, whether there is a risk premium, and, if there is one, whether it is positive or negative. Because of statistical problems, the empirical work in their paper does little to resolve the controversy. If there is no risk premium, we might conclude that the unbiased expectations theory holds, in which case the futures price is an unbiased estimator of the future spot price. When simple cash-and-carry arbitrage is difficult or costly, then the unbiased expectations theory may be appropriate.

Keynes (1930) assumed that there are two types of participants in futures markets: speculators and hedgers. He argued that if hedgers are net short, then speculators must be net long. But speculators will not go long unless the futures price is expected to rise, that is, unless $F < E(S_T)$. Keynes called the situation where the futures price is less than the expected spot price at delivery (and hence the futures price is expected to rise) normal backwardation. Put another way, the futures price is a downward biased estimator of the future expected spot price, and it reflects the existence of a positive risk premium: $E(S_T) - F > 0$.

Hicks (1939) reversed Keynes's theory by pointing out that there are situations in which hedgers are net long. In this case, called contango, speculators must be net short, and since they will not go short unless futures prices are expected to fall, it must be the case that $F > E(S_T)$. When markets are in contango, futures prices are expected to decline, and the futures price is an upward-biased estimator of the future expected spot price. The risk premium, $E(S_T) - F$, is negative under contango.

Regardless of whether markets are in normal backwardation or contango, speculators will trade only if they expect to be compensated for bearing risk—that is, the futures price contains a risk premium that could be positive or negative, and it can fluctuate and change signs over time, as long and short hedging activities vary. Telser (1960) first pointed out that hedging activities can be seasonal. Thus, there could be times that hedgers are net short and other times that they could be net long.

Stock index futures are interesting because they almost surely are always in normal backwardation. We know that stocks are risky assets, and if investors are risk averse, stocks are priced to provide a risk premium over and above the riskless interest rate. Thus $E(S_T) = S_0(1 + r + R\text{P})$, where r is the holding-period riskless rate and RP is the equity risk premium. But assuming (for simplicity) that the stocks pay no dividends, cash-and-carry arbitrage requires that $F = S_0(1 + r)$ (for discrete interest rates, and where r is an unannualized holding period riskless interest rate). Thus, $F < E(S_T)$ for futures and forward contracts on stocks and stock indices.

ELECTRICITY FORWARDS AND FUTURES

The issue of how electricity is priced in spot and forward wholesale power markets has become one of the most controversial topics facing utilities, power producers, regulators, political officials, accounting firms, and a broad array of financial market participants.

—Longstaff and Wang (2004, p. 1877)

The chapter closes with a discussion of nonstorable underlying assets, such as electricity. It is particularly challenging to estimate what the forward price for electricity should be.

Electricity futures contracts began trading on the New York Mercantile Exchange (NYMEX) in 1996, which now offers contracts with monthly delivery at several geographic locations. All of the contracts are cash settled. A bilateral forward market for electricity has existed for many years; recently, intermediaries often have brokered the deals between the different parties that want to lock in the forward price for electricity.

A generalized cost of carry futures/forwards pricing model is

$$F = Se^{(r-y)T}$$

where $F =$ theoretical futures/forward price that precludes arbitrage
$S =$ spot price of the underlying asset
$r =$ carrying costs, including interest and storage costs
$y =$ carry return from dividends (if a stock), coupons (if a bond), foreign interest (if a currency), leasing income (if a commodity that can be leased), or convenience yield
$T =$ time until delivery date

The cost of carry pricing model works well under six assumptions:

1. There are no transaction costs.
2. There are no taxes.
3. Proceeds from selling short are received and can be invested.
4. Borrowing rates equal lending rates.
5. The carry return is known (dividend amounts and timing usually must be estimated).
6. The underlying asset can be stored.

Concerning this last assumption, at one time it was widely believed that the underlying asset had to be storable in order to have futures and forward contracts on it. If the underlying asset cannot be stored, then cash-and-carry arbitrage, the force behind the basic forward/futures pricing models presented above, no longer works. To estimate a theoretical futures/forward price, we must reconsider the expectations model, which requires the estimation of two variables that are very difficult to estimate:

1. The expected spot price of the underlying asset at delivery.
2. An appropriate risk-adjusted discount rate that is applied to the expected spot price.

Electricity represents the ultimate in a nonstorable commodity; once it is generated, electricity must either be consumed or wasted. The spot price of electricity is determined by supply and demand at that moment and at that geographic location. There was a week in 1997 during which the average price of electricity at a Midwest location was $61/megawatts, but during that same week it traded as high as $7,500/megawatts. Peak intraday prices can be over 100 times higher than prices that existed just a few hours earlier and that will exist just a few hours subsequently.

Recognizing that nonstorability of electricity negates the typical cost of carry futures/forwards pricing model, Bessembinder and Lemmon (2002) developed an equilibrium electricity forward pricing model that predicts whether the actual forward price will be above or below the expected future spot price. The forward price is determined by the actions of electricity producers and retailers. The authors ignored the impact of outside speculators who may be trading futures contracts. The forward premium (the difference between the spot price expected at delivery and today's forward price) and forward discount represent compensation for bearing price and/or demand risk for electricity.

Their model predicts that the forward price will be above the expected future spot price (contango) during periods that spot prices possess a positively skewed distribution (e.g., during periods of peak daytime summer demand). During these periods, there is high demand for long forward positions to hedge against the risk of electricity price spikes. As a result, the forward price will rise above the expected future spot price.

If there is no skew to the distribution of spot electricity prices (primarily during the spring and fall, when electricity demand is low and demand risk is low), then they conclude that forward prices will be below expected spot prices (normal backwardation).

Bessembinder and Lemmon (2002) find support for their hypotheses. Further support is provided by Longstaff and Wang (2004), who examine over two years of "day-ahead" forward contracts for the existence of forward premia and discounts, depending on demand conditions. Estimating the expected price for delivery at a given hour tomorrow should be easier than estimating it for delivery a month hence. Many electricity forward contracts call for delivery in as short a time as an hour.

Other electricity forward/futures pricing models exist. Bühler and Müller-Merbach (2007) classify them into three categories: econometric models, reduced form models, and equilibrium models. Econometric models use historical and fundamental data to estimate what the futures price should be. Reduced form models specify a few risk factors, which are modeled as diffusion processes or jump processes. Bühler and Müller-Merbach extend Bessembinder and Lemmon's equilibrium model to derive the term structure of electricity forward prices.

Regardless of the model, we conclude that the cost of carry pricing model is of no value when determining theoretical futures/forward prices for electricity. The models that have been developed for electricity forward/futures pricing are complex and also require the estimation of the future expected spot price.

CONCLUSION

The cost of carry futures pricing model does a very good job estimating the futures prices for financial underlying assets and gold. Financial underlying assets include stocks, currencies, and interest rates/bonds. Agricultural commodities often have a convenience yield, and this is very difficult to measure. Because of this fact, the cost of carry pricing model (borrow, buy the spot good, and sell futures) only establishes an upper bound for pricing via arbitrage. Reverse cash-and-carry arbitrage (sell the spot good, lend the proceeds, and buy futures) is not effective because the underlying asset will not or cannot be sold or sold short. The convenience yield is the nonobservable benefit for having the spot good in inventory. When the convenience yield exists, futures prices for nonfinancial underlying assets will be less than the price normally predicted by the cost of carry model.

The other major futures pricing model is the expectations model, which states that the futures price equals the expected spot price on the delivery date, plus or minus a risk premium.

The survey article by Chow, McAleer, and Sequeira (2000) provides a nice complement to this chapter.

ENDNOTES

1. If compounding is not continuous, the analogous notation for the future value of a dollar to be received at time T is $(1+r)T$.
2. If compounding is not continuous, the model is $F = S(1+r)T$.
3. More precisely, d is the future value of the dividend yield. In other words, you must account for the interest that can be earned on the dividends received from owning the stock.
4. The discrete model is

$$F = S\left(\frac{1+r}{1+f}\right)^{T}.$$

5. The nonmonetary benefits behind the convenience yield are not the observable and measurable monetary carry return benefits. Rather, they are typically the benefits for having the underlying asset needed for producing another product. The oil refiner needs crude oil as a raw material. A cereal manufacturer needs wheat and oats as an input to manufacture cereal.
6. If futures prices are below spot prices, and if they decline as the delivery date moves farther into the future, we say that we are experiencing an inverted futures pricing curve.
7. For example, see Boyle (1989); Gay and Manaster (1984, 1986); Hemler (1990); Henrard (2006); Hranaiova, Jarrow, and Tomek (2005); Kane and Marcus (1986a,b); and Pirrong, Kormendi, and Meguire (1994). Chance and Hemler (1993) provide a nice survey of the early work.
8. LIBOR, the basis for Eurodollar futures contract, uses a convention of 360 days per year.
9. Chicago Board of Trade. Available at http://cbot.com/cbot/pub/cont_detail/0,3206, 1526+14431,00.html.
10. See Dubofsky and Miller (2003), Chapter 9, for a detailed discussion concerning how to determine the cheapest to deliver treasury security.
11. For simplicity, we are ignoring the interest that must be applied to the initial accrued interest. If the futures contract's delivery day is nearby (three months or less), this is inconsequential.

REFERENCES

Bessembinder, H., and M. L. Lemmon 2002. "Equilibrium Pricing and Optimal Hedging in Electricity Forward Markets," *Journal of Finance* 57, no. 3 (June): 1347–1382.

Boyle, P. P. 1989. "The Quality Option and Timing Option in Futures Contracts," *Journal of Finance* 44, no. 1 (March): 101–114.

Brennan, M. J. 1958. "The Supply of Storage," *American Economic Review* 48, no. 1 (March): 50–72.

Bühler, W., and J. Müller-Merbach. 2007. "Dynamic Equilibrium Valuation of Electricity Futures," (April). Working paper available at SSRN: http://ssrn.com/abstract=982791.

Chance, D. M., and M. L. Hemler. 1993. "The Impact of Delivery Options on Futures Prices: A Survey," *Journal of Futures Markets* 13, no. 2 (April): 127–155.

Chang, C. W., and J. S. K. Chang. 1990. "Forward and Futures Prices: Evidence from the Foreign Exchange Markets," *Journal of Finance* 45, no. 4 (September): 1333–1336.

Chow, Y., M. McAleer, and J. M. Sequeira. 2000. "Pricing of Forward and Futures Contracts," *Journal of Economic Surveys* 14, no. 2 (April): 215–253.

Cootner, P. H. 1960. "Returns to Speculators: Telser vs. Keynes," *Journal of Political Economy* 68, no. 4 (August): 396–404.

Cornell, B., and M. R. Reinganum. 1981. "Forward and Futures Prices: Evidence from the Foreign Exchange Markets," *Journal of Finance* 36, no. 5 (December): 1035–1045.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1981. "The Relation between Forward Prices and Futures Prices," *Journal of Financial Economics* 9, 4 (December): 321–346.

Dubofsky, D. A., and T. W. Miller. 2003. *"Derivatives:* Valuation and Risk Management". New York: Oxford University Press.

Dusak, K. 1973. "Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premiums," *Journal of Political Economy* 81, no. 6 (December): 1387–1406.

Fama, E. F., and K. R. French. 1987. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *Journal of Business* 60, no. 1 (January): 55–73.

French, K.R. 1983. "A Comparison of Futures and Forward Prices", *Journal of Financial Economics* 12, no. 3 (November): 311–342.

Gay, G. D., and S. Manaster. 1984. "The Quality Option Implicit in Futures Contracts," *Journal of Financial Economics* 13, no. 3 (September): 353–370.

Gay, G. D., and S. Manaster. 1986. "Implicit Delivery Options and Optimal Delivery Strategies for Financial Futures Contracts," *Journal of Financial Economics* 16, no. 1 (May): 41–72.

Grinblatt, M., and N. Jegadeesh. 1996. "Relative Pricing of Eurodollar Futures and Forward Contracts," *Journal of Finance* 51, no. 4 (September): 1499–1522.

Hemler, M. L. 1990. "The Quality Delivery Option in Treasury Bond Futures Contracts," *Journal of Finance* 45, no. 5 (December): 1565–1586.

Henrard, M. 2006. "Bonds Futures and Their Options: More than the Cheapest-to-Deliver; Quality Option and Margining," *Journal of Fixed Income* 16, no. 2 (September): 62–75.

Hicks, J. 1939. *Value and Capital*. Cambridge: Oxford University Press.

Hranaiova, J., R. A. Jarrow, and W. Tomek. 2005. "Estimating the Value of Delivery Options in Futures Contracts," *Journal of Financial Research* 28, no. 3 (September): 363–383.

Jarrow, R. A., and G. S. Oldfield. 1981. "Forward Contracts and Futures Contracts," *Journal of Financial Economics* 9, no. 4 (December): 373–382.

Kaldor, N. 1939. "Speculation and Economic Stability," *Review of Economic Studies* 7, no. 1 (October): 1–27.

Kane, A., and A. Marcus. 1986a. "The Quality Option in the Treasury Bond Futures Market: An Empirical Assessment," *Journal of Futures Markets* 6, no. 2 (Summer): 231–248.

Kane, A., and A. Marcus. 1986b. "Valuation and Optimal Exercise of the Wild Card Option in the Treasury Bond Futures Market," *Journal of Finance* 41, no. 1 (March): 195–207.

Keynes, J. M. 1930. *Treatise on Money*, vol. 2. London: Macmillan.

Longstaff, F. A., and A. W. Wang. 2004. "Electricity Forward Prices: A High-Frequency Empirical Analysis," *Journal of Finance* 59, no. 4 (August): 1877–1900.

MacKinley, A.C., and K. Ramaswamy. 1988. "Index Futures Arbitrage and the Behavior of Stock Index Futures Prices," Review of Financial Studies 1, no. 2 (Summer): 137–158.

Park, H. Y., and A. H. Chen. 1985. "Differences between Futures and Forward Prices: A Further Investigation of the Marking-to Market Effects," *Journal of Futures Markets* 5, no. 1 (February): 77–88.

Pindyck, R. S. 2001. "The Dynamics of Commodity Spot and Futures Markets: A Primer," *Energy Journal* 22, no. 3 (July): 1–29.

Pirrong, S. C., R. Kormendi, and P. Meguire. 1994. "Multiple Delivery Points, Pricing Dynamics, and Hedging Effectiveness in Futures Markets for Spatial Commodities," *Journal of Futures Markets* 14, no. 5 (August): 545–574.

Telser, L. G. 1958. "Futures Trading and the Storage of Cotton and Wheat," *Journal of Political Economy* 66, no. 3 (June): 233–255.

Telser, L. G. 1960. "Returns to Speculators: Telser versus Keynes: A Reply," *Journal of Political Economy* 68, no. 4 (August): 404–415.

Working, H. 1948. "Theory of the Inverse Carrying Charge in Futures Markets," *Journal of Farm Economics* 30, no. 1 (February): 1–28.

Working, H. 1949. "The Theory of the Price of Storage," *American Economic Review* 39, no. 6 (December): 1254–1262.

ABOUT THE AUTHOR

David Dubofsky is professor of finance and associate dean for research at the University of Louisville. Previously, he was a visiting academic scholar for the Office of Economic Analysis at the U.S. Securities and Exchange Commission and professor and department chair at Virginia Commonwealth University, and also held endowed professorships at Texas A&M University and Seattle University. Professor Dubofsky has published over 25 articles in peer-reviewed academic journals including the *Journal of Finance,* the *Journal of Financial and Quantitative Analysis,* and the *Journal of Money, Credit and Banking.* He is the author of two textbooks, the most recent of which is *Derivatives: Valuation and Risk Management,* coauthored with Tom Miller and published by Oxford University Press. He is a CFA charter holder. He earned a PhD in finance from the University of Washington, an MBA from the University of Houston, and a BE in chemical engineering from City College of New York.

CHAPTER 26

The Black-Scholes Option Pricing Model

A. G. MALLIARIS
Walter F. Mullady Sr. Professor of Business Administration, School of Business Administration, Loyola University Chicago

INTRODUCTION

Earlier chapters have introduced the reader to options as derivative securities. One reason for the phenomenal success of options is the fact that they can be priced. Pricing usually means that there is an active and liquid market with many buyers and sellers that generate a continuous flow of prices for a specific instrument. However, the pricing of options goes beyond supply and demand fundamentals. Market participants follow pricing methodologies that can be used as benchmarks. This is an attractive property for options because the presence of pricing methodologies allows for more effective arbitrage strategies.

Actually, when one compares the difficulties of pricing an underlying asset such as a stock, because of the uncertainty of the stream of future earnings and also the uncertainty of future discount rates, in comparison, the pricing of a European option is "relatively" easier than the pricing of its underlying asset. The relative easiness of pricing European options comes from the fact that puts and calls depend on six inputs. These inputs are the price of the underlying asset denoted by S, the striking price K, the time to maturity T, the volatility of the underlying σ, the interest rate r, and the distribution of dividends d, if applicable. For example, we can write the price of a European call as

$$c = f(S, K, T, \sigma, r, d) \tag{26.1}$$

And similarly we can write the price of a European put as

$$p = f(S, K, T, \sigma, r, d) \tag{26.2}$$

Furthermore, both the call and put in Equations 26.1 and 26.2 are related by the put-call parity expressed as

$$p + S = c + K \times e^{-rT} \tag{26.3}$$

Furthermore, financial reasoning allows one to propose that the partials of Equations 26.1 and 26.2 satisfy these relationships:

$$\frac{\partial c}{\partial S} > 0;\ \frac{\partial c}{\partial K} < 0;\ \frac{\partial c}{\partial \sigma} > 0;\ \frac{\partial c}{\partial T} > 0;\ \frac{\partial c}{\partial r} > 0;\ \frac{\partial c}{\partial d} < 0 \tag{26.4}$$

$$\frac{\partial p}{\partial S} < 0;\ \frac{\partial p}{\partial K} > 0;\ \frac{\partial p}{\partial \sigma} > 0;\ \frac{\partial p}{\partial T} > 0;\ \frac{\partial p}{\partial r} < 0;\ \frac{\partial p}{\partial d} > 0 \tag{26.5}$$

There are different approaches to pricing options: One may use arbitrage methods, tree diagrams, or the Black-Scholes option pricing model. These methods are interrelated. This chapter describes the fundamentals of Black-Scholes option pricing.

BRIEF HISTORY

Prior to Fischer Black and Myron Scholes (1973) several economists and researchers had attempted to derive a closed-form solution for Equation 26.1. For example, in 1877, Charles Castelli wrote a book titled *The Theory of Options in Stocks and Shares*. Castelli's book introduced the public to the hedging and speculation aspects of options but was not successful in pricing these securities. The mathematician Louis Bachelier (1900) offered the earliest known analytical valuation for options in his mathematics dissertation *Theorie de la Speculation* at the Sorbonne written under the supervision of the famous mathematician called Henri Poincaré. This thesis, originally written in French, has been translated into English and published in Paul Cootner's (1964) book.

Bachelier (1900) introduced the concept of a continuous random walk process that was a great innovation but also had a major problem since such a process generated share prices that allowed both negative security prices and option prices that exceeded the price of the underlying asset. Events such as the World War I, the stock market crash of 1929, and World War II diverted attention away from the pricing of options until the mid-1960s, when Paul Samuelson (1965) revisited the pricing of options. During that same year, Richard Kruizenga (1965), one of Samuelson's students, cited Bachelier's work in his dissertation. In 1964, James Boness's dissertation focused on options. In his work, Boness developed a pricing model that made a significant theoretical advance. Finally, almost concurrently and independently in the early 1970s, Robert Merton (1973) writing his Ph.D. thesis under Paul Samuelson at MIT and Fischer Black and Myron Scholes as young researchers discovered the now-famous Black Scholes (1973) model, also known as the Black-Scholes-Merton model.

BLACK-SCHOLES FORMULA

The closed-form solution obtained by Black and Scholes (1973) is:

$$c = S \times N(d_1) - K \times e^{-rT} \times N(d_2) \tag{26.6}$$

where c = theoretical call premium
S = current stock price
T = time until option expiration
K = option striking price
r = risk-free interest rate
N = cumulative standard normal distribution
e = exponential term (2.7183)

$$d_1 = \frac{\ell n\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right) \times T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

σ = annualized standard deviation of stock returns
ℓn = natural logarithm

As stated, our goal is to help the reader understand how Equation 26.6 is obtained. Before we present the derivation of the formula, we wish to emphasize that the equation has a very appealing financial interpretation. Equation 26.6 says that the price of a call option is the difference between the expected value of the underlying stock and the expected cost of the option. The expected value of the stock price and the expected cost of the option involve probabilities that are associated with the standardized normal distribution that is used to describe the continuous random walk component of the stock price.

To familiarize the reader with the Black-Scholes model, we first discuss the assumptions made in the development of the model, its mathematical derivations, and its financial reasoning.

ASSUMPTIONS OF THE BLACK-SCHOLES MODEL

The key assumptions of the Black-Scholes model are:

- The price of the underlying instrument S_t follows a geometric Brownian motion with constant drift μ and constant volatility σ:

$$dS_t = \mu \times S_t \times dt + \sigma \times S_t \times dZ_t \tag{26.7}$$

- It is possible to short sell the underlying stock.
- There are no limits to arbitrage.
- Trading in all securities is continuous.
- There are no transaction costs or taxes.
- All securities are perfectly divisible (e.g., it is possible to buy 1/100th of a share).
- It is possible to borrow and lend cash at a constant risk-free interest rate.
- The stock does not pay a dividend.

Next we evaluate these assumptions.

DISCUSSION OF ASSUMPTIONS

In financial economics, as in science, assumptions are made to simplify the complexity of the problem to be solved. If the assumptions are too strong, then the problem may be trivialized. Thus, a careful assessment of the appropriateness of these assumptions is necessary. Merton (1975, 1982) offers a lengthy discussion of the appropriateness of these assumptions, and serious students are encouraged to read these two papers. Essentially, short selling, no arbitrage limitations, continuous trading with no transactions costs or taxes, and perfect divisibility of all securities are standard assumptions that have been used in solving several other problems, such as portfolio selection. Also, the assumption of borrowing and lending at the constant risk-free interest rate and the assumption of no dividends can be modified easily.

From the list of assumptions, the most significant one is Equation 26.7. Actually until that equation was proposed as the appropriate equation to describe the behavior of returns of the underlying asset, no researcher had succeeded in pricing options correctly because they were all considering special cases of Equation 26.7. For example, Bachelier's work assumed the special case of the equation with $\mu = 0$ and $\sigma = 1$.

ITO PROCESS

In this section, we explain Equation 26.7, technically called as an Ito process, since it plays a very important role in pricing derivative securities. A detailed explanation of the mathematical and financial meaning of this equation can be found in Malliaris and Brock (1982).

Mathematically, any continuous function can be approximated with a polynomial function of an appropriate degree. However, most asset prices are very random and require a high-order polynomial representation. Thus, polynomial approximations are not useful in describing asset returns. In addition, polynomials of high order are very complex.

An alternative approach to polynomial approximation is to model share price differences as a random walk, that, is let $S_t - S_{t-1} = dZ_t$. This assumes the asset price has neither a trend nor volatility other than the randomness in the continuous random walk. We can generalize and assume that asset returns have a drift $\mu \neq 0$ and volatility $\sigma > 0$ that allows us to write Equation 26.7. This says that returns of any asset written as $\frac{dS_t}{S_t}$ can be expressed as the sum of a constant return per trading interval dt and a random product of two factors, the volatility σ of the asset and a random shock dZ.

Taking the expectation of Equation 26.7, we get

$$E\left(\frac{dS_t}{S_t}\right) = E(\mu) \times dt + E(\sigma dZ_t) \tag{26.8}$$

where $E(\mu) = \mu$ since μ is a constant

$E(\sigma dZ_t) = \sigma E(dZ_t) = 0$ since $E(dZ_t) = 0$

$Var(dZ_t) = 1$

In other words, dZ_t as a random variable describing continuous random walk follows a standardized normal distribution with mean zero and variance 1.

The assumption by Black and Scholes (1973) and Merton (1973) that the underlying asset follows an Ito process as in Equation 26.7 was an extremely successful step that allowed the researchers to use an already existing mathematical theory with financial reasoning to solve the pricing of derivatives.

Assuming Equation 26.7 means that the price of the underlying asset follows the stochastic process given by

$$S_t = S_o \times e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma Z_t} \tag{26.9}$$

where as in Equation 26.7

$$\begin{aligned} \mu &= \text{drift of the asset} \\ \sigma &= \text{volatility of the asset} \\ Z_t \text{ with } t \in [0, \infty) &= \text{a Wiener process with independent increments} \\ &\quad \text{that are normally distributed with mean zero} \\ &\quad \text{and variance 1} \end{aligned}$$

Taking the logarithm of Equation 26.9, we get

$$\ell nS_t = \ell nS_o + \left(\mu - \frac{\sigma^2}{2}\right) \times t + \sigma \times Z_t \tag{26.10}$$

This last equation says that the stock price described by Equation 26.9 is distributed log normally with expectation and variance given by

$$E\,(\ell nS_t) = \ell nS_o + \left(\mu - \frac{\sigma^2}{2}\right) \times t \tag{26.11}$$

$$Var\,(\ell nS_t) = \sigma^2 \times t \tag{26.12}$$

EXAMPLE

To help the reader appreciate the usefulness of Equation 26.9, which is the solution of Equation 26.7, we give the next example. Let the IBM stock price today be $S_o = \$100$, suppose $\mu = 0.15, \sigma = 0.10$. Where do you expect IBM to be a year from now? Calculations show:

$$\begin{aligned} E\,(\ell nS_t) &= \ell n\,(100) + \left(0.15 - \frac{0.10^2}{2}\right) \times 1 = 4.75\ \mathit{Std.\ Deviation} = \sigma \times \sqrt{t} \\ &= 0.10 \times \sqrt{1} = 0.10 e^{(4.75 - 2\times 0.10)} < S(1\ \mathit{year}) < e^{(4.75+2\times 0.10)}\ 94.63 \\ &< S(1\ \mathit{year}) < 141.18 S(t) = S(0) \times e^{\mu t} S(1\ \mathit{year}) = 100 \times e^{(0.15)\times 1} = 116.18 \end{aligned}$$

In words, the IBM stock price is expected to be at \$116.18 a year from today with 95 percent probability that the price will be no lower than \$94.63 or higher than \$141.18.

EXCEL APPLICATION

Here we give an intuitive description of (26.9) with reference to Exhibit 26.1. In this table we collect from Yahoo.com 40 recent daily closing IBM prices. These prices are also illustrated graphically in Exhibit 26.2.

For modeling purposes, the Black-Scholes equation requires a mathematical expression for prices, such as that shown in Exhibits 26.1 and 26.2. How can we check to see if the prices in these exhibit follow an Ito process? In column 3 of Exhibit 26.1, we compute the daily IBM returns and calculate the average historical return of the 40 daily returns. This daily average is 0.0010504, which, annualized becomes $0.0010504 \times 250 = 0.2626$. The volatility of these 40 returns is computed as the annualized standard deviation of daily returns and is computed as $0.0102656 \times \sqrt{250} = 0.162313$. Both these calculations assume 250 trading days per year. To complete checking that the prices in Exhibit 26.1 satisfy Equation 26.7, we solve for dZ. Since these prices are daily, we use the daily constant average return and constant daily volatility to obtain daily dZs in column 4 of Exhibit 26.1 using

$$dZ_i = \frac{r_i - \mu_{ri}}{\sigma_{ri}}$$

and also compute the mean and variance of the dZ in the last column. Recall that earlier we postulated that $E\,(dZ) = 0$ and $Var\,(dZ) = 1$ is assumed. Indeed, the mean and variance of the dZ values in Exhibit 26.1 are as postulated.

Note that if we were to graph the distribution of IBM stock prices, it would follow a lognormal distribution while its returns would follow a normal distribution. Furthermore, the daily dZs, as an approximation to the continuous random walk dZ of Equation 26.7, also are normally distributed with mean zero and variance 1. Exhibits 26.3 and 26.4 are frequency approximations to a normal distribution and illustrate the distributions of returns and distributions of dZs of Exhibit 26.1.

SIMPLE DERIVATION OF BLACK-SCHOLES

In this section, we present financial reasoning along with computations from mathematical integration to illustrate the derivation of the Black-Scholes model.

Consider a call on asset $S(t)$ following Equation 26.7 with strike price K and time T to expiration. Today, at time t, we expect at expiration either $S(T) > K$ or $S(T) \leq K$ with corresponding probabilities

$$p = \text{Prob}\,[S(T) > K] \tag{26.13}$$

$$1 - p = \text{Prob}\,[S(T) \leq K] \tag{26.14}$$

Exhibit 26.1 Daily Price Data for IBM

Day	Price	r_i	dZ
1	113.49		
2	113.13	−0.0031770	−0.4118140
3	112.98	−0.0013270	−0.2315680
4	110.73	−0.0201160	−2.0618710
5	112.64	0.0171021	1.5636331
6	112.71	0.0006213	−0.0418040
7	112.05	−0.0058730	−0.6744200
8	111.23	−0.0073450	−0.8178230
9	109.69	−0.0139420	−1.4604390
10	110.90	0.0109707	0.9663589
11	109.22	−0.0152650	−1.5892930
12	109.04	−0.0016490	−0.2629950
13	110.00	0.0087656	0.7515538
14	111.45	0.0130957	1.1733608
15	113.24	0.0159334	1.4497886
16	113.44	0.0017646	0.0695719
17	112.00	−0.0127750	−1.3467840
18	114.57	0.0226871	2.1076845
19	115.37	0.0069584	0.5755088
20	116.69	0.0113765	1.0058891
21	118.19	0.0127727	1.1418926
22	117.88	−0.0026260	−0.3581600
23	117.62	−0.0022080	−0.3174160
24	115.55	−0.0177560	−1.8319530
25	115.80	0.0021612	0.1082082
26	117.35	0.0132964	1.1929079
27	116.00	−0.0115710	−1.2294550
28	115.95	−0.0004310	−0.1443190
29	115.13	−0.0070970	−0.7936720
30	114.52	−0.0053120	−0.6198200
31	116.63	0.0182571	1.6761409
32	116.67	0.0003429	−0.0689190
33	116.86	0.0016272	0.0561873
34	116.78	−0.0006850	−0.1690320
35	116.25	−0.0045490	−0.5454300
36	116.51	0.0022341	0.1153031
37	117.30	0.0067576	0.5559566
38	117.71	0.0034892	0.2375708
39	117.80	0.0007643	−0.0278700
40	119.03	0.0103873	0.9095287
41	118.36	−0.0056450	−0.6521890

$\mu_{r_i} = 0.0010504$
$\sigma_{r_i} = 0.0102656$
$\mu_{dz} = 0$
$\sigma_{dz} = 1$

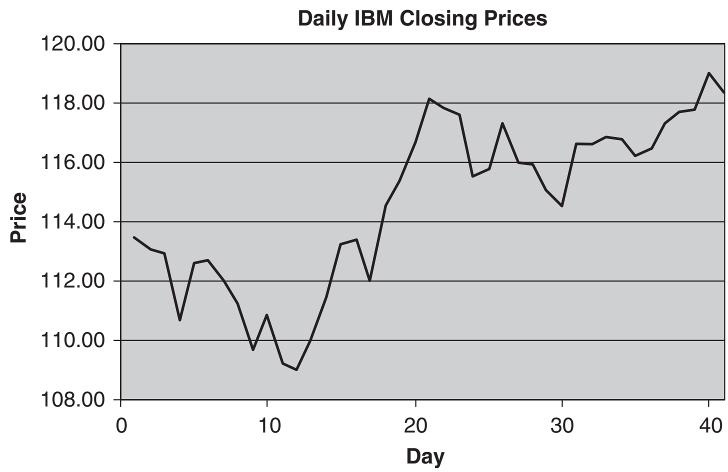


Exhibit 26.2 Daily IBM Closing Prices

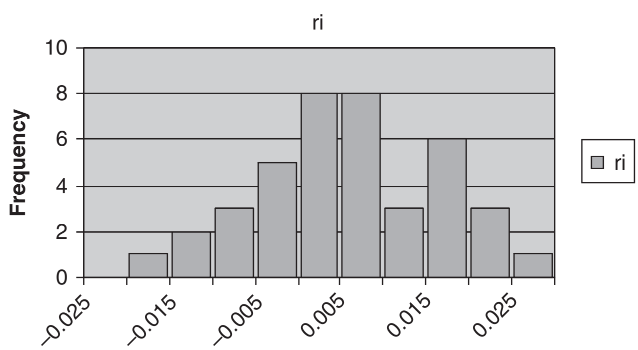


Exhibit 26.3 Frequency Distribution of Returns

Corresponding to Equations 26.13 and 26.14, the price of a (European or American since there are no dividends) call at expiration, denoted $c(T)$, is

$$c(T) = E\{\max[0, S(T) - K]\} \tag{26.15}$$

and discounted to t, Equation 26.15 becomes

$$c(t) = e^{-rT} \times E\{\max[0, S(T) - K]\} \tag{26.16}$$

Combine Equations 26.13 and 26.14 with Equation 26.15 to derive the price of a call today, $c(t)$ as

$$\begin{aligned} c(t) &= p \times e^{-rT} \times E[S(T) - K] + (1 - p) \times 0 \\ &= p \times e^{-rT} E[S(T)|S(T) > K] - p \times e^{-rT} \times K \end{aligned} \tag{26.17}$$

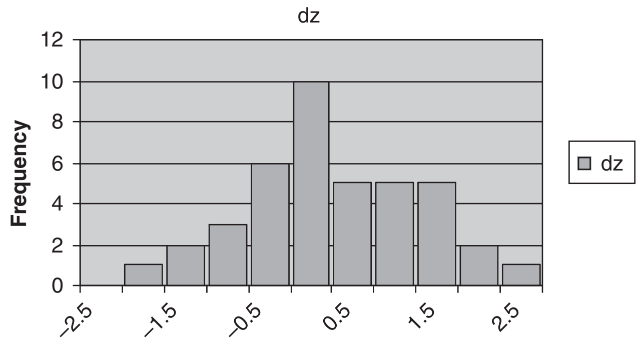


Exhibit 26.4 Frequency Distribution of Errors

Calculations in Hull (2006) show that

$$p = P\left[S(T) > K\right] = N\left[\frac{\ell\text{n}\dfrac{S(t)}{K} + \left(r - \dfrac{\sigma^2}{2}\right) \times T}{\sigma\sqrt{T}}\right]$$
$$= N\left[d_2\right] \tag{26.18}$$

For ease of notation

$$d_2 \equiv \frac{\ell\text{n}\dfrac{S(t)}{K} + \left(r - \dfrac{\sigma^2}{2}\right) \times T}{\sigma\sqrt{T}} \tag{26.19}$$

We also follow Hull (2006) to conclude

$$p \times E\left[S(t)|S(t) > K\right] = S(t) \times e^{rT} \times N\left[\frac{\ell\text{n}\dfrac{S}{K} + \left(r + \dfrac{\sigma^2}{2}\right) \times T}{\sigma\sqrt{T}}\right]$$
$$= S(t) \times e^{rT} \times N\left[d_1\right] \tag{26.20}$$

As in Equation 26.19,

$$d_1 \equiv \frac{\ell\text{n}\dfrac{S}{K} + \left(r + \dfrac{\sigma^2}{2}\right) \times T}{\sigma\sqrt{T}} \tag{26.21}$$

Using Equations 26.18 and 26.20 in Equation 26.17, we get

$$\begin{aligned} c(t) &= e^{-rT} \times p \times E[S(T)|S(T) > K] - e^{-rT} \times K \times p \\ &= e^{-rT} \times S(t) \times e^{rT} \times N[d_1] - e^{-rT} \times K \times N[d_2] \\ &= S(t) \times N[d_1] - e^{-rT} \times K \times N[d_2] \end{aligned} \tag{26.22}$$

In other words, the Black-Scholes model in Equation 26.22 is the result of certain simplifying assumptions, financial reasoning, and the computation of three integrals.

NUMERICAL EXAMPLE

Let the current stock price of IBM be \$100. Assume as earlier that its volatility is expected to be $\sigma = 0.15$. Compute the one year at-the-money call and put price with strike price \$100, $r = 5$ percent, and no dividends. Check also for the put-call parity.

$$\begin{aligned} d_1 &= \frac{\ell n\left(\dfrac{100}{100}\right) + \left(0.05 + \dfrac{(0.15)^2}{2}\right) \times 1}{0.15\sqrt{1}} \\ &= \frac{0 + (0.05 + 0.01125)}{0.15} \\ &= 0.4083 \end{aligned}$$

$$\begin{aligned} d_2 &= \frac{\ell n\left(\dfrac{100}{100}\right) + \left(0.05 - \dfrac{(0.15)^2}{2}\right) \times 1}{0.15\sqrt{1}} \\ &= \frac{0 + (0.05 - 0.01125)}{0.15} \\ &= 0.2583 \end{aligned}$$

Compute the probabilities

$$\begin{aligned} N(d_1) &= N(0.4083) \\ &= 0.658471 \\ N(d_2) &= N(-0.9571) \\ &= 0.601937 \end{aligned}$$

Conclude that the call and put prices are

$$\begin{aligned} c(t) &= S(t) \times N(d_1) - e^{-rT} \times K \times N(d_2) \\ &= 100 \times 0.658471 - e^{-(0.05)\times(1)} \times 100 \times 0.601937 \\ &= 8.59 \end{aligned}$$

Put price

$$N(-d_1) = N(-0.4083) = 0.341529$$

$$N(-d_2) = N(-0.2583) = 0.398063$$

$$\begin{aligned} p &= e^{-rT} \times K \times N(-d_2) - S(t) \times N(-d_1) \\ &= e^{-(0.05)\times 1} \times 100 \times 0.398063 - 100 \times 0.341529 \\ &= 3.71 \end{aligned}$$

We check to the put-call parity

$$\begin{aligned} p + S &= 3.71 + 100 = 103.71 \\ c + K \times e^{-rT} &= 8.59 + 100 \times e^{-(0.05)\times 1} = 8.59 + 95.12 \\ &= 103.71 \end{aligned}$$

Thus, put-call parity $p + S = c + K \times e^{-rT}$ holds.

THE GREEKS

From Equation 26.22, we can obtain several partial derivatives and also certain useful higher-order partials as expressions of the responsiveness of the call (and put) prices to changes in certain inputs.

Delta

Delta is a measure of the sensitivity the calculated option value has to small changes in the share price. Delta for a European call is given by:

$$\frac{\partial c}{\partial S} = N(d_1)$$

Gamma

Gamma is a measure of the calculated delta's sensitivity to small changes in share price. Gamma for a European call is given by:

$$\frac{\partial^2 c}{\partial S^2} = \frac{N'(d_1)}{S \times \sigma\sqrt{T}}$$

where $N'(d_1)$ here and below is given by $N'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Theta

Theta measures the calculated option value's sensitivity to small changes in time till maturity. Theta for a European call is given by:

$$\frac{\partial c}{\partial T} = \frac{S \times N'(d_1) \times \sigma}{2\sqrt{T}} - r \times K \times e^{-rT} \times N(d_2)$$

Vega

Vega measures the calculated option value's sensitivity to small changes in volatility. Vega for a European call is given by:

$$\frac{\partial c}{\partial \sigma} = S \times \sqrt{T} \times N'(d_1)$$

Rho

Rho measures the calculated option values of the portfolio with respect to the interest rate. Rho for a European call is given by:

$$\frac{\partial c}{\partial r} = K \times T \times e^{-rT} \times N(d_2)$$

All partials for puts and calls with dividends can be found in Hull (2006, pp. 360–362).

RISK-NEUTRAL PRICING

To illustrate the importance of continuous arbitrage and pricing under conditions of risk neutrality, consider the nominal value of a portfolio $P(t)$ consisting of the underlying stock $S(t)$ and its call $c(t)$:

$$P(t) = N_1(t) \times S(t) + N_2(t) \times c(t) \tag{26.23}$$

where $N_1(t) =$ number of shares
$N_2(t) =$ number of calls

As before, we assume that

$$dS(t) = \mu \times S(t) \times dt + \sigma \times S(t) \times dZ \tag{26.24}$$

In view of Equation 26.24, we can compute changes in the price of the call option $dc(t)$ using Taylor's theorem or Ito's lemma in Malliaris and Brock (1982):

$$\begin{aligned} dc(t) &= \frac{\partial c}{\partial t}dt + \frac{\partial c}{\partial S}dS + \frac{1}{2} \times \frac{\partial^2 c}{\partial S^2} \times \sigma^2 \times S^2 \times dt \\ &= \left(\frac{\partial c}{\partial t} + \frac{1}{2} \times \frac{\partial^2 c}{\partial S^2} \times \sigma^2 \times S^2\right) dt + \frac{\partial c}{\partial S}dS \end{aligned} \tag{26.25}$$

Using Equations 26.24 and 26.25, which each describes the infinitesimal changes in the values of the underlying stock $S(t)$ and its call option $c(t)$, we can track changes in the investor's portfolio.

Next

$$\begin{aligned} dP &= N_1 \times dS + N_2 \times dc \\ &= N_1 \times dS + N_2 \left[\left(\frac{\partial c}{\partial t} + \frac{1}{2} \times \frac{\partial^2 c}{\partial S^2} \times \sigma^2 \times S^2\right) dt + \frac{\partial c}{\partial S}dS\right] \end{aligned} \tag{26.26}$$

This last equation says that the portfolio change is stochastic or random because of dS in both the first and last term.

Suppose that

$$\frac{N_1}{N_2} = -\frac{\partial c}{\partial S} \tag{26.27}$$

This means that the portfolio is continuously hedged so it is not exposed to risk or randomness. This can be accomplished because randomness as described by the continuous random walk in dZ is "nice" and because continuous costless trading is assumed to keep up with this nice randomness. Using Equations 26.27 and 26.26, conclude

$$\begin{aligned} dP &= -N_2 \times \frac{\partial c}{\partial S} dS + N_2 \left[\left(\frac{\partial c}{\partial S} + \frac{1}{2} \times \frac{\partial^2 c}{\partial S^2} \times \sigma^2 \times S^2 \right) dt + \frac{\partial c}{\partial S} dS \right] \\ &= N_2 \left(\frac{\partial c}{\partial t} + \frac{1}{2} \times \frac{\partial^2 c}{\partial S^2} \times \sigma^2 \times S^2 \right) dt \end{aligned} \tag{26.28}$$

Let $N_2 = 1$ for simplicity. Since we can trade continuously to offset risk evolving as a continuous random walk, the return of this continuously portfolio must equal the risk-free return r. This can be written as

$$\frac{dP}{P} = r \times dt \tag{26.29}$$

This last equation can be rewritten as

$$\frac{\left(\frac{\partial c}{\partial t} + \frac{1}{2} \times \frac{\partial^2 c}{\partial S^2} \times \sigma^2 \times S^2 \right)}{-\frac{\partial c}{\partial S} \times S + c} = r \times dt$$

which, upon rearrangement, becomes the famous parabolic partial differential equation of the Black-Scholes model with boundary conditions

$$\begin{aligned} c\,(t,S) &= 0 \\ c\,(T,S) &= \max\,[0,\, S - K] \end{aligned}$$

This equation has as its solution the Black-Scholes equation given earlier and repeated once more here

$$c\,(S,K,T,\sigma,r) = S \times N\,(d_1) - K \times e^{-rT} \times N\,(d_2)$$

This analysis also can be formulated using advanced mathematical methods. Chalamandaris and Malliaris (2008) give a detailed presentation of advanced mathematical methods in deriving the Black-Scholes model.

CONCLUSION

This chapter has introduced the reader to the Black-Scholes-Merton model by identifying its assumptions and illustrating its mathematical derivation using intuitive financial reasoning. Numerical examples also have been presented to help the reader understand practical aspects of this celebrated model. The analytical power of the Black-Scholes-Merton model comes from the brilliant assumption that the returns of the underlying asset follow an Ito process. This assumption allowed financial theorists to use financial reasoning with an extensive inventory of mathematical techniques to solve successfully for the pricing of contingent claims. Unlike many other scientific discoveries that are not often easily modified, the Black-Scholes-Merton model has been extended and adapted to numerous underlying assets, thus offering pricing solutions as benchmark prices. This in turn encouraged the development and implementation of numerous trading strategies that involve hedging, speculation, and arbitrage. The truly phenomenal increase in trading volume in derivatives can be accounted for, at least to some extent, by the availability of the Black-Scholes-Merton pricing methodologies and their current use in risk management.

REFERENCES

Bachelier, L. 1900. "Theory of Speculation,", in P. H. Cootner, ed., *The Random Character of Stock Market Prices*, chap. 2. Cambridge, MA: MIT Press, 1964.

Black, M., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, no. 3 (May/June): 637–654.

Boness, A. J. 1964. "Elements of a Theory of Stock-Option Value," *Journal of Political Economy* 72, no. 2 (April): 163–175.

Chalamandaris, G., and A.G. Malliaris. 2008. "Ito's Calculus and the Black-Scholes Options Pricing Model," in C. F. Lee and A. Lee, eds., *Handbook of Quantitative Finance*. New York: Springer.

Cootner, P. H. ed. 1964. *The Random Character of Stock Market Prices*, Cambridge, MA: MIT Press.

Hull, J. C. 2006. *Options, Futures, and Other Derivatives*, 6th ed. Upper Saddle River, NJ: Prentice-Hall.

Kruizenga, R. J. 1965. *Put and Call Options: A Theoretical and Market Analysis*. PhD diss., MIT.

Malliaris, A. G., and W. A. Brock. 1982. *Stochastic Methods in Economics and Finance*. Amsterdam: North-Holland.

McKean Jr., Henry, P. 1965. "Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem of Mathematical Economics," *Industrial Management Review* 6, no. 2 (Spring): 32–39.

Merton, R. C. 1973. "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4, no. 1 (Spring): 141–183.

Merton, R. C. 1975. "Theory of Finance from the Perspective of Continuous Time," *Journal of Financial and Quantitative Analysis* 10, no. 4: 659–674.

Merton, R. C. 1982. "On the Mathematics and Economic Assumptions of Continuous-Time Models," in W. F. Sharpe and C. M. Cootner, eds., *Financial Economics: Essays in Honor of Paul Cootner*. Amsterdam: North-Holland.

Samuelson, P. A. 1965. "Rational Theory of Warrant Pricing" *Industrial Management Review* 6, no. 2 (Spring): 13–31.

ABOUT THE AUTHOR

A. G. Malliaris is professor of economics and finance at Loyola University Chicago and holds the Walter F. Mullady Sr. Chair in Business Administration. He graduated from the Athens School of Economics and Business with a BA in economics and earned a PhD in economics from the University of Oklahoma and a second PhD in mathematics from the University of Chicago. He has coauthored with William A. Brock two books, *Stochastic Methods in Economics and Finance* and *Differential Equations, Stability and Chaos in Dynamic Economics.* He has also authored two other books, *Foundations of Futures Markets* (1999) and *Economic Uncertainty, Instabilities and Asset Bubbles* (2005). He specializes in financial economics and has made several contributions in the area of futures and options markets.

CHAPTER 27

The Black-Scholes Legacy

Closed-Form Option Pricing Models

ANTÓNIO CÂMARA
Assistant Professor of Finance, Watson Family Chair in Commodity and Financial Risk Management, Spears School of Business, Oklahoma State University

INTRODUCTION

The Black-Scholes (1973) closed-form option pricing model lies at the core of financial economics, and it is certainly the most important formula in derivatives. According to Rubinstein (1994), "the [Black-Scholes] model is widely viewed as one of the most successful in the social sciences and [is] perhaps (including its binomial extension) the most widely used formula, with embedded probabilities in human history." The Black-Scholes model is consistent with general equilibrium in the economy, and it is an arbitrage-free system of valuation equations.

There are several closed-form option pricing models that were obtained prior to Black and Scholes. The most popular of these closed-form option pricing formulas are the models derived by Bachelier (1900); Sprenkle (1961); Ayres (1963); Boness (1964); Samuelson (1965); Baumol, Malkiel, and Quandt (1966); and Chen (1970). These models are not obtained in general equilibrium economies or arbitrage-free economies, and, therefore, they do not consider the time value of money and the adjustment of risky payoffs according to modern finance theory. However, like the Black-Scholes model, most of these models already assume that the stock price follows a geometric Brownian motion and, therefore, that it has a lognormal distribution at the end of each period. The exception is Bachelier, who assumes that the stock price follows an arithmetic Brownian motion and, therefore, that it has a normal distribution at the end of each period.

This chapter reviews some of the closed-form pricing models that followed the Black-Scholes model. A closed-form option pricing model is an option pricing formula that can be evaluated in a finite number of standard operations that involve well-known functions and gives exact price for the assumptions of the model. Since what is a "standard operation" and "well-known function" is open to debate, there might be disagreement among researchers and practitioners about what is a closed-form option pricing model. A numerical solution, since it provides only an approximation to the option price, is not a closed-form pricing model. Option pricing formulas that depend on infinite sums are not closed-form option pricing models because they do not entail a finite number of standard operations.

In this chapter, we review closed-form European call and put option pricing models that are obtained using either arbitrage arguments or equilibrium arguments. Goldenberg (1991) provides a method to obtain closed-form option pricing equations using arbitrage arguments for a class of diffusion processes. Camara (2003) presents a paradigm to derive closed-form option pricing equations using equilibrium arguments for the transformed-normal class of distributions. Haug (1998) reviews many closed-form solutions obtained under the assumption that the underlying asset value has a lognormal distribution.

First we discuss the Black-Scholes closed-form solution. Then we present closed-form option pricing models that assume a single lognormal underlying variable, followed by closed-form solutions for option prices that depend on two or more underlying lognormal variables. Next we turn to closed-form option pricing models when there is a single nonlognormal underlying variables followed by models where there are several underlying stochastic variables and at least one of these is not lognormal. Exhibit 27.1 summarizes the main models of the four generations.

THE BLACK-SCHOLES MODEL

This section discusses the closed-form solutions of the Black-Scholes call and put option pricing models. We discuss several methods that can be used to derive the Black-Scholes closed-form call and put option pricing models. The first method is the arbitrage-free option pricing method proposed by Black and Scholes (1973), who write: "If options are correctly priced in the market, it should not be possible to make sure profits by creating portfolios of long and short positions in options and their underlying stocks. Using this principle, a theoretical valuation formula for options is derived" (p. 637).

Black and Scholes derive their closed-form option pricing models making seven explicit assumptions:

1. The risk-free interest rate, r, is constant.
2. Borrowing and lending at the risk-free interest rate has no restrictions.[1]
3. The stock price follows a geometric Brownian motion through time and, therefore, has a lognormal distribution at the end of each period. The variance of log returns is constant.
4. The stock pays no dividends.
5. The option is an European-style option and, therefore, can be exercised only at maturity.[2]
6. There are no transaction costs.[3]
7. Short-selling is allowed.[4]

According to assumption (7c), the stock price S follows a geometric Brownian motion:

$$dS = S\mu dt + S\sigma dW \tag{27.1}$$

Exhibit 27.1 Classification of Closed-Form Option Pricing Models

Precursors of Black and Scholes:
- Bachelier (1900)
- Sprenkle (1961)
- Ayres (1963)
- Boness (1964)
- Samuelson (1965)
- Baumol, Malkiel, and Quandt (1966)
- Chen (1970)

The Black and Scholes model is the first model to adjust correctly the risky payoff of the option for time value of money and risk.

First Generation: A Single Lognormal Underlying
- Merton (1973): Option on a stock that pays continuous dividends.
- Black (1976): Option on a futures contract.
- Garman and Kohlhagen (1983): Foreign exchange option.

Second Generation: Two Lognormal Underlyings
- Merton (1973): Option on a stock with random interest rates.
- Geske (1978): Option on a stock with stochastic dividends.
- Stapleton and Subrahmanyam (1984): Option with stochastic strike price.
- Grabbe (1983): Foreign exchange option with random interest rates.

Third Generation: A Single Non-lognormal Underlying
- Merton (1976): Ruin option pricing model with a jump-diffusion.
- Brennan (1979): Options on normal cash flows.
- Rubinstein (1983): Options on a displaced lognormal firm that has both risky and riskless assets.
- Camara (2003): Options on the Su system of distributions.

Fourth Generation: Two Underlyings where One Is Non-lognormal
- Heston (1993a): Options with stochastic volatility.

where μ = instantaneous stock mean return
σ = instantaneous stock volatility
dW = a Wiener process

The Wiener or Brownian motion process has a normal distribution with mean zero and variance dt, that is, $dW \sim N(0, dt)$. Equation 27.1 means that the stock return $\frac{dS}{S} = \frac{S_t - S_{t-1}}{S_{t-1}}$ over a small period of time $dt = t - (t-1)$ is given by its mean return μ plus its volatility σ times a normal random variable.

As pointed by Black and Scholes, the option price F depends only on the stock price S and time t, that is, $F = F(S, t)$. Hence, the option price follows the stochastic process:

$$dF = F_s dS + F_t dt + \frac{1}{2} F_{ss} S^2 \sigma^2 \, dt \tag{27.2}$$

Black and Scholes construct a portfolio of a long position on a stock and a short position on $1/F_s$ options. Hence the instantaneous return on this portfolio is:

$$d\Pi = dS - \frac{1}{F_s}dF \tag{27.3}$$

$$= -\frac{1}{F_s}\left(F_t + \frac{1}{2}F_{ss}S^2\sigma^2\right)dt, \tag{27.4}$$

where Equation 27.4 follows by substituting Equation 27.2 into Equation 27.3.

As we can see, Equation 27.4 does not depend on any random component. Therefore, the return of this portfolio is certain. This portfolio is a hedged portfolio. Therefore, the return on the initial investment $S - 1/F_sF$ is certain and instantaneously equal to:

$$d\Pi = (S - 1/F_sF)r\,dt \tag{27.5}$$

Equating the return in Equation 27.5 to the return of Equation 27.4 yields the Black-Scholes (1973) partial differential equation (PDE):

$$rSF_s - rF + F_t + \frac{1}{2}F_{ss}S^2\sigma^2 = 0 \tag{27.6}$$

The Black-Scholes PDE (Equation 27.6) can be used to price any derivative in an arbitrage-free economy when assumptions 1 to 7 are satisfied. Different derivatives have different boundary conditions for the PDE (Equation 27.6) yielding different option pricing formulas. Black and Scholes derive closed-form option pricing models for call and put options. The boundary condition for the call (put) option is its final payoff $Max(S - K, 0)$ ($Max(K - S, 0)$), where K is the strike price. Solving Equation 27.6 subject to the appropriate boundary condition, using the heat-transfer equation of physics, Black and Scholes derive their closed-form option pricing models:

$$P_c = S_0N(d_1) - Ke^{-rT}N(d_2) \tag{27.7}$$

$$P_p = Ke^{-rT}N(-d_2) - S_0N(-d_1) \tag{27.8}$$

where

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

$d_2 = d_1 - \sigma\sqrt{T}$

$N(.)$ = cumulative distribution function of a standard normal random variable

P_c = call price

P_p = put price

T = time to maturity of the options

As we can see, the Black-Scholes formula depends on the stock price S_0, the strike price K, the stock volatility σ, the risk-free interest rate r, and the time to maturity of the option T. All these parameters, with the exception of the stock volatility, are observable variables. Exhibit 27.2 illustrates how the Black-Scholes call option price changes when the stock volatility changes. As we can see from the figure, when the volatility increases, the option price also increases. The intuition behind this result is that the option has an upside potential but no downside risk.

Black and Scholes, in the same article, also show that the closed-form option pricing formulas 27.7 and 27.8 can be obtained using the capital asset pricing model (CAPM), which is an equilibrium model of asset prices. Equations 27.7 and 27.8 are important because, as highlighted by Merton (1973), "an exact formula for an asset price, based on observable variables only, is a rare finding from a general equilibrium model (p 161)." However, it was Rubinstein (1976) who later showed that the no-arbitrage model of Black and Scholes could be obtained in a general equilibrium economy.

The solution to the stochastic differential equation (SDE) 27.1 is given by:

$$S_T = exp\left(ln(S_0) + \mu T - \frac{\sigma^2}{2}T + \sigma W(T)\right) \tag{27.9}$$

Taking logarithms in Equation 27.9, and since $E(W(T)) = 0$ and $Var(W(T)) = T$, we can easily see that the stock price at date T, S_T, as a lognormal distribution:

$$S_T \sim \Lambda\left(ln(S_0) + \mu T - \frac{\sigma^2}{2}T, \sigma^2 T\right) \tag{27.10}$$

where Λ = lognormal distribution

Rubinstein (1976) derives the Black-Scholes call and put option pricing formulas 27.7 and 27.8 by assuming that the stock price has a lognormal distribution (Equation 27.10), aggregate consumption has a lognormal distribution (Equation 27.11), and the representative agent has a power utility function of consumption

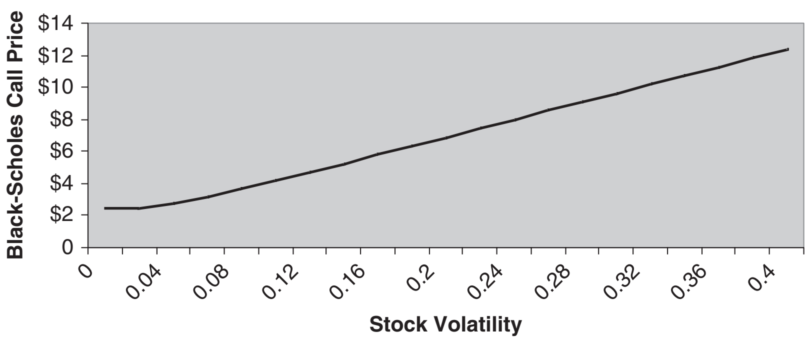


Exhibit 27.2 The Impact of Risk on Option Prices
This exhibit assumes that the stock price is 100, the strike price is 100, the interest rate is 5 percent, and the call option has a maturity of 6 months.

(Equation 27.12):

$$C_T \sim \Lambda\left(ln(C_0) + \mu_c T - \frac{\sigma_c^2}{2}T, \sigma_c^2 T\right) \qquad (27.11)$$

$$U_t(C_t) = d^t C_t^{1-b}/(1-b) \qquad (27.12)$$

where μ_c = expected growth rate of consumption
σ_c = consumption volatility
d = time discount factor
b = coefficient of proportional risk aversion

In a discrete-time, continuous state-space economy, it is not possible to create and maintain a dynamic hedge portfolio or to derive the Black-Scholes PDE (Equation 27.6). Rubinstein (1976) assumes that there is a representative agent that maximizes expected utility of consumption when it takes its optimal consumption and investment decisions. With this assumption, it is possible to obtain equilibrium pricing equations for the bond, the stock, and the call:

$$B(0,T) = E\left[d^T\left(\frac{C_T}{C_0}\right)^{-b} \times 1\right] \qquad (27.13)$$

$$S_0 = E\left[d^T\left(\frac{C_T}{C_0}\right)^{-b} \times S_T\right] \qquad (27.14)$$

$$P_c = E\left[d^T\left(\frac{C_T}{C_0}\right)^{-b} \times Max(S_T - K, 0)\right] \qquad (27.15)$$

Equations 27.13 to 27.15 are the equilibrium pricing equations, also known as the Euler equations. The Black-Scholes call price (Equation 27.7) is obtained in two steps.

1. We evaluate Equations 27.13 to 27.15 using the distributions (Equations 27.10 and 27.11) the utility function (Equation 27.12), and noting that $Cov(S_T, C_T) = \sigma_{cs}T$. The three resulting equations will depend on preference parameters.
2. We use the resulting equations of the bond and the stock in the resulting equation of the call to eliminate the preference parameters from this equation. The final result is the Black-Scholes call valuation (Equation 27.7) obtained in a general equilibrium model. It is a well-known result in financial economics that if there is equilibrium then there is no arbitrage.

FIRST GENERATION OF MODELS (ONE LOGNORMAL UNDERLYING)

We include in the first generation of closed-form option pricing models all those formulas that relax a single assumption of the Black-Scholes economy while maintaining the assumption of a lognormal underlying variable. Among these models, we include option pricing models of options on stocks that pay a continuous dividend yield, futures options, and foreign exchange options.

Merton (1973) generalizes the Black-Scholes option pricing equations (27.7 and 27.8) by considering a stock that pays a continuous dividend yield q.[5] The stock price still follows the stochastic differential equation (Equation 27.1). The owner of the stock receives dividends in the dollar amount of qS during the infinitesimal period of time dt. Following Black and Scholes (1973), we can construct a portfolio of a long position on a stock and a short position on $1/F_s$ options. In this case, the instantaneous return on this portfolio is not given by Equation 27.3 since the holder of the portfolio is also entitled to receive dividends. The instantaneous return on this portfolio is:

$$d\Pi = dS - \frac{1}{F_s}dF + qSdt \tag{27.16}$$

The remainder of the analysis to derive the PDE for options on stocks that pay a continuous dividend yield is identical to the analysis developed by Black and Scholes. In this case, the PDE is given by:

$$(r-q)\,SF_s - rF + F_t + \frac{1}{2}F_{ss}S^2\sigma^2 = 0 \tag{27.17}$$

As we can see, the PDE for options on stocks that pay a continuous dividend yield q (Equation 27.17) differs from the Black-Scholes PDE (Equation 27.6) only in their first term. The riskless return r that appeared in the first term of the Black-Scholes PDE was replaced by $r-q$ in this PDE. The boundary conditions for the call and put options written on a stock remain unchanged, and the prices of these options are now given by the following equations:

$$P_c = S_0e^{-qT}\,N(d_1) - Ke^{-rT}\,N(d_2) \tag{27.18}$$

$$P_p = Ke^{-rT}N(-d_2) - S_0e^{-qT}N(-d_1) \tag{27.19}$$

where:

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

Black (1976) generalizes the Black-Scholes option pricing equations (Equations 27.7 and 27.8) by considering options written on a futures contract. The price of the futures contract still follows the stochastic differential equation (Equation 27.1), where S stands now for futures price. Following Black and Scholes (1973), we can construct a portfolio of a long position on a futures contract and a short position on $1/F_s$ options. In this case, the instantaneous return on this portfolio is still given by Equations 27.3 and 27.4. However, since there is no payment when we enter into the futures contract, the initial investment in the futures contract is zero. Hence from Equation 27.5, the return on the initial investment $S - 1/F_s F$, with $S = 0$, is certain and instantaneously equal to:

$$d\Pi = -1/F_s F r\, dt \tag{27.20}$$

The remainder of the analysis to derive the PDE for options on futures is identical to the analysis developed by Black and Scholes. In this case, the PDE for options on futures is given by:

$$-rF + F_t + \frac{1}{2}F_{ss}S^2\sigma^2 = 0 \tag{27.21}$$

As we can see, this is the Black-Scholes PDE without the first term. The formulas of call and put options written on futures contracts are given by:

$$P_c = e^{-rT}[S_0 N(d_1) - K N(d_2)] \tag{27.22}$$

$$P_p = e^{-rT}[K N(-d_2) - S_0 N(-d_1)] \tag{27.23}$$

where:

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

and S_0 in this model is the futures price at date 0 for delivery at date T.

Garman and Kohlhagen (1983) generalize the Black-Scholes option pricing equations (Equations 27.7 and 27.8) by considering foreign exchange options. The spot price of the deliverable currency (domestic units per foreign unit) still follows the geometric Brownian motion (Equation 27.1). Here, the interest rate r stands for domestic interest rate while r_f stands for foreign interest rate.

The PDE for foreign exchange options is given by:

$$(r - r_f)SF_s - rF + F_t + \frac{1}{2}F_{ss}S^2\sigma^2 = 0 \tag{27.24}$$

As we can see, this PDE for foreign exchange options is identical to the PDE for options on a stock that pays a continuous dividend yield (Equation 27.17) with r_f replacing q. Then the closed-form solutions of foreign exchange options are:

$$P_c = S_0 e^{-r_f T} N(d_1) - K e^{-rT} N(d_2) \tag{27.25}$$

$$P_p = K e^{-rT} N(-d_2) - S_0 e^{-r_f T} N(-d_1) \tag{27.26}$$

where:

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r - r_f + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

which are identical to Equations 27.18 and 27.19 with r_f replacing q.

SECOND GENERATION OF MODELS (TWO LOGNORMAL UNDERLYINGS)

The second generation of closed-form option pricing formulas assumes that there is more than a single stochastic underlying variable and that all the underlying stochastic variables are lognormally distributed. These formulas relax a specific assumption of the Black-Scholes model.

Merton (1973) extends the Black-Scholes model (Equations 27.7 and 27.8) to stochastic interest rates and uses nonarbitrage arguments to obtain a PDE that values derivatives. The value of the option in this model depends on the stock price S, the bond price B, and t. Hence the two-dimensional Ito's Lemma is applied to derive the process followed by the option. Merton constructs a zero investment portfolio with the stock, the bond, and the option. Noting that this portfolio has a zero return, in order to avoid arbitrage opportunities to arise in the economy, it is possible to derive a PDE that generalizes Equation 27.6. Merton changes variables and actually solves a PDE that is a special case of Equation 27.6. This PDE is solved subject to the appropriate boundary conditions in order to obtain closed-form option pricing equations for the call and put options. The stock price in his model follows the geometric Brownian motion (Equation 27.1), and the zero coupon bond price with face value \$1 has these dynamics:

$$dB = \alpha B + \delta B dZ \tag{27.27}$$

where $dZ =$ Wiener process correlated with dW, $dZdW = \rho dt$

The closed-form solutions obtained by Merton (1973) are:

$$P_c = S_0 N(d_1) - KB(0,T)N(d_2) \tag{27.28}$$

$$P_p = K\,B(0,T)N(-d_2) - S_0 N(-d_1) \tag{27.29}$$

where

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) - ln(B(0,T)) + \frac{V^2}{2}T}{V\sqrt{T}}$$

$$d_2 = d_1 - V\sqrt{T}$$

$$V^2 = \frac{1}{T}\int_0^T (\sigma^2(s) + \delta^2(s) - 2\rho\sigma\delta(s))\,ds$$

$B(0,T) =$ price at date 0 of the bond that matures at date T

Geske (1978) obtains closed-form solutions for call and put options written on a stock that pays a stochastic dividend yield. Since dynamic trade on the dividend yield is not possible, Geske (1978) uses the equilibrium approach of Rubinstein (1976). It is assumed that the representative agent has a power utility of consumption (Equation 27.12) and that the stock price, aggregate consumption, and the (inverse of the) dividend yield are lognormally distributed, as in Equations 27.10, 27.11, and:

$$q^{-1} \sim \Lambda\left(ln(\mu_{q^{-1}})T - \frac{\sigma_q^2}{2}T, \sigma_q^2 T\right) \tag{27.30}$$

Geske (1978) obtains the following closed-form solutions for the prices of call and put options:

$$P_c = S_0 \Phi N(d_1) - Ke^{-rT} N(d_2) \tag{27.31}$$

$$P_p = Ke^{-rT} N(-d_2) - S_0 \Phi N(-d_1) \tag{27.32}$$

where

$$d_1 = \frac{ln\left(\frac{S_0\Phi}{K}\right) + \left(r + \frac{\sigma_v^2}{2}\right)T}{\sigma_v\sqrt{T}}$$

$$d_2 = d_1 - \sigma_v\sqrt{T}$$

$$\Phi = \mu_{q^{-1}} \exp\left[(-\sigma_{sq} + b\sigma_{cq})T\right]$$

$$\sigma_v^2 = \sigma^2 + \sigma_q^2 - 2\sigma_{sq}$$

σ_{sq} = covariance between the stock and the dividend yield

σ_{cq} = covariance between aggregate consumption and the dividend yield

The formulas derived by Geske (1978) are obtained following the same steps of Rubinstein to obtain the Black and Scholes model (1973). The stock price in Equation 27.14 takes into account total returns since the holder of the stock receives dividends. Hence, when Equation 27.14 is evaluated, the resulting equation depends on one risk premium, the stock risk premium, $-b\sigma_{sc}$. The stock price of Equation 27.15 considers only capital gains since the price of the option is not adjusted for dividends. Hence, when Equation 27.15 is evaluated, the resulting equation depends on two risk premiums, the stock risk premium, $-b\sigma_{sc}$, and the risk premium of the dividend yield, $-b\sigma_{cq}$. When the resulting equilibrium price equation (27.14) is substituted in the resulting pricing equation (27.15), the stock risk premium is eliminated from the option price equation. This explains why the stock risk premium $-b\sigma_{sc}$ does not appear in Equations 27.31 and 27.32 but the risk premium of the dividend yield $-b\sigma_{cq}$ affects those two pricing equations.

Stapleton and Subrahmanyam (1984) extend the methodology of Rubinstein (1976) to obtain closed-form expressions for the prices of call and put options written on a stock when the strike price K is stochastic. First, they derive a risk-neutral valuation relationship (RNVR) (i.e., they show that they can assume actual distributions for the underlying variables and that they can use the risk-neutral distributions when pricing derivatives written on those underlying variables). Their setting is identical to Rubinstein's since there is no dynamic trading. A representative agent who maximizes his expected utility of consumption derives prices when he selects his optimal portfolio and investment. Then they use this RNVR in order to obtain closed-form solutions for the prices of call and put options written on a stock when the strike price is stochastic. Since preferences do not affect the risk-neutral distributions, this explains why neither the stock risk premium nor the risk premium of the strike price affects their formulas.

Stapleton and Subrahmanyam (1984) assume that the exercise price has a lognormal distribution:

$$K \sim \Lambda\left(ln(K_0) + \mu_K T - \frac{\sigma_K^2}{2}T, \sigma_K^2 T\right) \tag{27.33}$$

where $K_0 =$ strike price at date 0
$\mu_K =$ expected strike price
$\sigma_K =$ volatility of the strike price

Stapleton and Subrahmanyam also assume that the stock price and consumption are lognormal as in Equations 27.10 and 27.11, and that the utility function of the representative is a power utility function, as in Equation 27.12. Their closed-form option pricing formulas are:

$$P_c = S_0 N(d_1) - K_0 e^{-rT} N(d_2) \tag{27.34}$$

$$P_p = K_0 e^{-rT} N(-d_2) - S_0 N(-d_1) \tag{27.35}$$

where

$$d_1 = \frac{ln\left(\frac{S_0}{K_0}\right) + \left(r + \frac{\sigma_\alpha^2}{2}\right)T}{\sigma_\alpha \sqrt{T}}$$

$$d_2 = d_1 - \sigma_\alpha \sqrt{T}$$

$$\sigma_\alpha^2 = \sigma^2 + \sigma_K^2 - 2\sigma_{sK}$$

$\sigma_{sK} =$ covariance between S and K

The second generation of closed-form option pricing models also includes, among others, the formulas of Grabbe (1983) and Amin and Jarrow (1991). Grabbe links the work of Merton (1973) on stochastic interest rates to the work of Garman and Kohlhagen on foreign exchange options to derive closed-form option pricing equations when the spot price of the deliverable currency, the domestic interest rate, and the foreign interest rate follow three different correlated geometric Brownian motions. As remarked by Amin and Jarrow the pricing approach of Grabbe does not integrate a full-fledged term structure model into the valuation framework. Amin and Jarrow extend the analysis and derive further closed-form solutions on call options using the Heath, Jarrow, and Morton (1992) model of term structure. This approach has the advantage of considering explicitly a continuum of traded domestic and foreign bonds into the analysis.

THIRD GENERATION OF MODELS (ONE NONLOGNORMAL UNDERLYING)

This section presents closed-form option pricing models for cases where the option is written on a single underlying variable, which has a nonlognormal distribution at the maturity of the option.

Merton (1976) investigates the pricing of options by assuming that the stock price follows a jump-diffusion process:

$$dS = [\mu - \lambda(Y - 1)]Sdt + \sigma SdW + (Y - 1)Sd\pi \tag{27.36}$$

where μ = expected stock return conditional on the nonoccurrence of jumps
σ = stock volatility
λ = mean number of jumps per unit of time
dW = Wiener process as in the Black-Scholes model
$d\pi$ = Poisson process, which has value 1 with probability λdt or value 0 with probability $1 - \lambda dt$
$Y - 1$ = percentage change in the stock price if a jump occurs

It is assumed that dW and $d\pi$ are independent.

Merton (1976) derives explicit option pricing equations for call and put options when the stock price follows a jump-diffusion process as in Equation 27.36 and the size of the jumps Y are lognormally distributed. However, these explicit pricing equations are not closed-form expressions since the price depends on an infinite sum. Nevertheless, Merton derives closed-form option pricing equations for the special case where there is a positive probability of immediate ruin. In this case, if a jump occurs, $Y = 0$ and the rate of return of the stock if the jump occurs is $Y - 1 = -100$ percent. The jump-diffusion process of this special case is obtained by substituting $Y = 0$ into Equation 27.36, yielding:

$$dS = (\mu + \lambda)Sdt + \sigma SdW - Sd\pi \tag{27.37}$$

When the stock price follows the jump-diffusion process 27.37, the call option written on that stock must satisfy the following partial differential equation:

$$(r + \lambda)SF_s - (r + \lambda)F + F_t + \frac{1}{2}F_{ss}S^2\sigma^2 = 0 \tag{27.38}$$

Equation 27.38 is the Black-Scholes partial differential equation (Equation 27.6) where their interest rate r is replaced by $r + \lambda$. Hence the closed-form pricing equation of the call in Merton's model is the Black-Scholes call pricing equation, where their interest rate r is replaced by $r + \lambda$. In Merton's model, the put must satisfy a different partial differential equation. However, using the call-put parity, $P_p = S_0 - P_c - Ke^{-rT}$, it is possible to derive the closed-form solution for the price of the put. These pricing equations are given by:

$$P_c = S_0N(d_1) - Ke^{-(r+\lambda)T}N(d_2) \tag{27.39}$$

$$P_p = Ke^{-rT}(1 - e^{-\lambda T}) + Ke^{-(r+\lambda)T}N(-d_2) - S_0N(-d_1) \tag{27.40}$$

where

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r + \lambda + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

Brennan (1979) extends the ideas of Rubinstein (1976) and introduces the concept of RNVR. By definition, a RNVR is an explicit option pricing equation derived in a discrete-time, continuous state-space economy, which does not depend on preference parameters. Equilibrium prices in a RNVR are obtained such that the asset return under this RNVR is the risk-free rate of return. Brennan prices options in a single-period economy, where the Euler equation of the call is given by:

$$P_c = E\left[\frac{U_T'(C_T)}{E[U_T'(C_T)]} Max(S_T - K, 0)\right] \tag{27.41}$$

where $U_T'(C_T) =$ marginal utility function

Brennan (1979) assumes that the representative agent has preferences that display constant absolute risk aversion (CARA) and, in particular, that the representative agent has a negative exponential utility function of consumption:

$$U_T(C_T) = \frac{1}{\alpha} exp(\alpha C_T) \tag{27.42}$$

where α is a preference parameter and that aggregate consumption has a normal distribution:

$$C_T \sim N(C_0 + \mu_c T, \sigma_c^2 T), \tag{27.43}$$

and that the underlying variable, a cash flow, has a normal distribution:

$$S_T \sim N(S_0 + \mu T, \sigma^2 T) \tag{27.44}$$

The closed-form pricing equations of the Brennan's (1979) normal RNVR are given by:

$$P_c = e^{-rT} \sigma \sqrt{T} n(d_1) + (S_0 - Ke^{-rT})\, N(-d_1) \tag{27.45}$$

$$P_p = (Ke^{-rT} - S_0)\, N(d_1) + e^{-rT} \sigma \sqrt{T} n(d_1) \tag{27.46}$$

where $d_1 = \dfrac{K - S_0 e^{rT}}{\sigma \sqrt{T}}$

$n(.) =$ standard normal density function

Rubinstein (1983) generalizes the analysis of Black-Scholes (1973) by assuming that the underlying variable, a stock or a firm, follows a displaced geometric Brownian motion. If the underlying asset follows a displaced geometric Brownian motion, then it has, at the end of each moment, a displaced lognormal distribution:

$$S_T \sim \Lambda\left(\beta, \mu T - \frac{\sigma^2}{2} T, \sigma^2 T\right) \tag{27.47}$$

where $\beta =$ threshold or lower bound of the underlying asset

The actual displaced lognormal has the same shape of the actual standard lognormal, but it is displaced by its threshold β. While the standard lognormal has a lower bound at 0, the displaced lognormal has a lower bound at β. If the underlying asset is a firm, then β represents the riskless assets of the firm at date T. Such firms have both riskless and risky assets. This contrasts with the Black-Scholes model applied to the firm value. When the underlying variable of the Black-Scholes model is a firm, such a firm only has risky assets. This firm does not have riskless assets.[6] The closed-form solutions written on an underlying asset that has a displaced-lognormal distribution are given by:

$$P_c = (S_0 - \beta e^{-rT})\, N(d_1) - e^{-rT}(K - \beta)\, N(d_2) \tag{27.48}$$

$$P_p = (K - \beta)\, e^{-rT} N(-d_2) - (S_0 - \beta e^{-rT})\, N(-d_1) \tag{27.49}$$

where
$$d_1 = \frac{ln\left(\dfrac{S_0 e^{rT} - \beta}{K - \beta}\right) + \dfrac{\sigma^2}{2}T}{\sigma\sqrt{T}}$$
$$d_2 = d_1 - \sigma\sqrt{T}$$

Camara (2003) extends the RNVR analysis of Brennan (1979) to derive new closed-form solutions for the prices of call and put options in a single-period economy where the Euler equation of the call is given by Equation 27.41. Camara assumes that the representative agent has a power utility function (Equation 27.12), that aggregate consumption has a lognormal distribution (Equation 27.11), and that the underlying asset value, S_T, has a S_U system of distributions of Johnson (1949), that is,

$$Sinh^{-1}(S_T) \sim N\,(\mu T, \sigma\sqrt{T}) \tag{27.50}$$

The call and the put option pricing equations of the Camara's (2003) model are given by:

$$\begin{aligned} P_c = {} & \frac{1}{2}\exp\left\{\frac{\sigma^2}{2}T - rT + sinh^{-1}\left[S_0\, exp\left(rT - \frac{\sigma^2}{2}T\right)\right]\right\} N(d_1) \\ & - \frac{1}{2}\exp\left\{\frac{\sigma^2}{2}T - rT - sinh^{-1}\left[S_0\, exp\left(rT - \frac{\sigma^2}{2}T\right)\right]\right\} N(d_2) \\ & - Ke^{-rT} N(d_3) \end{aligned} \tag{27.51}$$

$$\begin{aligned} P_p = {} & Ke^{-rT} N(-d_3) \\ & + \frac{1}{2}\exp\left\{\frac{\sigma^2}{2}T - rT - sinh^{-1}\left[S_0\, exp\left(rT - \frac{\sigma^2}{2}T\right)\right]\right\} N(-d_2) \\ & - \frac{1}{2}\exp\left\{\frac{\sigma^2}{2}T - rT + sinh^{-1}\left[S_0\, exp\left(rT - \frac{\sigma^2}{2}T\right)\right]\right\} N(-d_1) \end{aligned} \tag{27.52}$$

where
$$d_1 = \frac{ln\left[\frac{\sqrt{1+S_0^2\,exp(2rT-\sigma^2T)}+S_0\,exp\left(rT-\frac{\sigma^2}{2}T\right)}{\sqrt{1+K^2+K}}\right]+\sigma^2T}{\sigma\sqrt{T}}$$
$$d_2 = d_1 - 2\sigma\sqrt{T}$$
$$d_3 = d_2 + \sigma\sqrt{T}$$

There are several other closed-form solutions to price call and put options when there is a single stochastic underlying variable whose terminal distribution is not lognormal. Cox (1975) assumes that the stock price is governed by $dS = \mu S dt + \gamma S^{n/2} dW$ and derives an explicit formula for the price of a call that depends on an infinite sum. Schroder (1989) expresses the constant elasticity of variance (CEV) option pricing formula of Cox in terms of the noncentral chi-square distribution, which allows the application of well-known approximation formulas and the derivation of a closed form solution. Heston, Loewenstein, and Willard (2006) obtain a new formula for the CEV process.

Cox and Ross (1976) assume that the SDE of the stock price is $dS = \mu S dt + \sigma dW$ and use the Ornstein-Uhlenbeck process with an absorbing barrier at zero in order to derive a closed-form option pricing model. Lo and Wang (1995) derive a Black-Scholes closed-formula with an adjusted volatility parameter, adjusted to account for the negative autocorrelation in the trending Ornstein-Uhlenbeck process.

Heston (1993b) derives closed-form solutions for call and put options written on a stock that has a gamma distribution. His analysis extends that of Rubinstein (1976) and Brennan (1979). Similar methodology was used by Gerber and Shiu (1994), who derive closed-form solutions for option prices written on a stock that has an inverse Gaussian distribution. Camara and Heston (2008) extend the model of Merton (Equations 27.39 and 27.40) where there is a positive probability of immediate ruin, by deriving closed-form option pricing solutions when the stock price has big jumps downward and big jumps upward.

FOURTH GENERATION OF MODELS

This section discusses option pricing when there are two or more stochastic underlying variables and at least one of those variables is not lognormal.

Heston (1993a) discusses option pricing under stochastic volatility. It is assumed that the stock price follows the stochastic differential equation (Equation 27.1) with stochastic volatility. The stochastic process followed by the variance is:

$$d\sigma^2 = \kappa(\theta - \sigma^2)dt + \phi\sigma dZ, \tag{27.53}$$

where $dZdW = \rho_{s\sigma}\, dt$

In equilibrium, all assets satisfy this PDE:

$$\frac{1}{2}\sigma^2 S^2 F_{ss} + \rho_{s\sigma}\phi\sigma^2 S F_{s\sigma^2} + \frac{1}{2}\phi^2\sigma^2 F_{\sigma^2\sigma^2} + rSF_s + [\kappa(\theta - \sigma^2) - \lambda(S, \sigma^2, t)]\, F_{\sigma^2} - rF + F_t = 0 \tag{27.54}$$

This PDE depends on $\lambda(S, \sigma^2, t)$ (i.e., the market price of risk of volatility, which Heston (1993b) sets $\lambda(S, \sigma^2, t) = \lambda\sigma^2$). The price of the call is obtained by solving Equation 27.54 subject to the appropriate boundary conditions. The price of the call option is given by:

$$P_c = S_0 P_1 - KB(0, T) P_2 \tag{27.55}$$

where P_1 and P_2 = probabilities that are not available in closed form

These probabilities can be obtained using characteristic functions, which leads some researchers to claim that option prices are obtained in quasi–closed-form solutions rather than in closed-form solutions. Bakshi and Chen (1997) derive an option pricing formula that allows for stochastic interest rates, stochastic dividends, and stochastic volatility. This approach that yields quasi–closed-form solutions was extended by Duffie, Pan, and Singleton (2000) to affine jump-diffusions.

CONCLUSION

This chapter reviews and offers a new classification of the literature post–Black-Scholes on closed-form solutions for the pricing of European style call and put options. Existing closed-form solutions are classified according to the number of stochastic underlying variables and their final distribution. The first generation of models includes all closed-form option pricing equations where the option depends on a single lognormal underlying variable. The second generation of models includes all closed-form option pricing equations where the option depends on at least two lognormal underlying variables. The third generation of models includes all closed-form option pricing equations where the option depends on only a single non-lognormal underlying variable. The fourth generation of models includes all closed-form option pricing equations where the option depends on two or more stochastic underlying variables and at least one of these in nonlognormal.

ENDNOTES

1. Bergman (1995) studies option pricing with differential interest rates for borrowing and for lending but obtains option pricing bounds instead of a closed-form pricing model.
2. Roll (1977), Geske (1979), and Whaley (1981) derive a closed-form solution for the price of an American call option that matures at date T when there is a known dividend D paid at date t, $t < T$. Longstaff (1990) derives closed-form expressions for options that can be extended by the optionholder.
3. Leland (1985) investigates the valuation of options with transaction costs but obtains option pricing bounds instead of a closed-form pricing model.
4. Bergman (1995) also analyses the impact on option prices of a return differential between long and short positions in the stock.
5. See Merton (1973, footnote 62).
6. A related strand of the literature studies equity and debt as options written on the assets of the firm. See, for example, Leland (1994) and Toft and Prucyk (1997) who obtain several closed-form pricing equations.

REFERENCES

Amin, K., and R. Jarrow. 1991. "Pricing Foreign Currency Options under Stochastic Interest Rates," *Journal of International Money and Finance* 10: 310–329.

Ayres, H. 1963. "Risk Aversion in the Warrants Market," *Industrial Management Review* 4: 497–505.

Bachelier, L. 1900. "Theory of Speculation," in Cootner, P. ed., *The Random Character of Stock Market Prices* (1964). Cambridge; MA: MIT Press.

Bakshi, G., and Z. Chen. 1997. "An Alternative Valuation Model for Contingent Claims," *Journal of Financial Economics* 44: 123–165.

Baumol, W., B. Malkiel, and R. Quandt. 1966. "The Valuation of Convertible Securities," *Quarterly Journal of Economics* 80: 48–59.

Bergman, Y. 1995. "Option Pricing with Differential Interest Rates," *Review of Financial Studies* 8: 475–500.

Black, F. 1976. "The Pricing of Commodity Contracts," *Journal of Financial Economics* 3: 167–179.

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81: 637–659.

Boness, A. 1964. "Elements of a Theory of Stock Option Values," *Journal of Political Economy* 72: 163–175.

Brennan, M. 1979. "The Pricing of Contingent Claims in Discrete Time Models," *Journal of Finance* 34: 53–68.

Camara, A. 2003. "A Generalization of the Brennan-Rubinstein Approach for the Pricing of Derivatives," *Journal of Finance* 58: 805–820.

Camara, A., and S. Heston. 2008. "Closed-Form Option Pricing Formulas with Extreme Events," *Journal of Futures Markets* 28: 213–311.

Chen, A. 1970. "A Model of Warrant Pricing in a Dynamic Market," *Journal of Finance* 25: 1041–1060.

Cox, J. 1975. "Notes on Option Pricing I: Constant Elasticity of Variance Diffusions," *Mimeo*. Stanford University.

Cox, J., and S. Ross. 1976. "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics* 3: 145–166.

Duffie, D., J. Pan, and K. Singleton. 2000. "Transform Analysis and Asset Pricing for Affine Jump-Diffusion," *Econometrica* 68: 1343–1376.

Garman, M., and S. Kohlhagen. 1983. "Foreign Currency exchange Values," *Journal of International Money and Finance* 2: 231–237

Gerber, H., and E. Shiu. 1994. "Option Pricing by Esscher Transforms," *Transactions of Society of Actuaries* 46: 99–141

Geske, R. 1978. "The Pricing of Options with Stochastic Dividend Yield," *Journal of Finance* 33: 617–625.

Geske, R. 1979. "A Note on an Analytical Valuation Formula for Unproctected American Call Options on Stocks with Known Dividends," *Journal of Financial Economics* 7: 63–81.

Goldenberg, D. 1991. "A Unified Method for Pricing Options on Diffusion Processes," *Journal of Financial Economics* 29: 3–34.

Grabbe, J. 1983. "The Pricing of Call and Put Options on Foreign Exchange," *Journal of International Money and Finance* 2: 239–253.

Haug, E. 1998. *The Complete Guide to Option Pricing Formulas*. New York: McGraw-Hill.

Heath, D., R. Jarrow, and A. Morton. 1992. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," *Econometrica* 60: 77–105.

Heston, S. 1993a. "A Closed-Form Solution of Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies* 6: 327–343.

Heston, S. 1993b. "Invisible Parameters in Option Prices," *Journal of Finance* 48: 933–947.

Heston, S., M. Loewenstein, and G. Willard. 2006. "Options and Bubbles," *Review of Financial Studies* 20: 359–390.
Johnson, N. 1949. "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika* 36: 149–176.
Leland, H. 1985. "Option Pricing with Transaction Costs," *Journal of Finance* 40: 1283–1301.
Leland, H. 1994. "Corporate Debt Value, Bond covenants, and Optimal Capital Structure," *Journal of Finance* 49: 1213–1252.
Lo, A., and J. Wang. 1995. "Implementing Option Pricing Models When Asset Returns Are Predictable," *Journal of Finance* 50: 87–129.
Longstaff, F. 1990. "Pricing Options with Extendible Maturities: Analysis and Applications," *Journal of Finance* 45: 935–957.
Merton, R. 1973. "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4: 141–183.
Merton, R. 1976. "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics* 3: 125–144.
Roll, R. 1977. "An Analytical Valuation Formula for Unprotected American Call Options on Stocks with Known Dividends," *Journal of Financial Economics* 5: 251–258.
Rubinstein, M. 1976. "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics and Management Science* 7: 407–425.
Rubinstein, M. 1983. "Displaced Diffusion Option Pricing," *Journal of Finance* 38: 213–217.
Rubinstein, M. 1994. "Implied Binomial Tree," *Journal of Finance* 49: 771–818.
Samuelson, P. 1965. "Rational Theory of Warrant Pricing," *Industrial Management Review* 6: 13–31.
Schroder, M. 1989. "Computing the Constant Elasticity of Variance Option Pricing Formula," *Journal of Finance* 44: 211–219.
Sprenkle, C. 1961. "Warrant Prices as Indicators of Expectations," *Yale Economic Essays* 1: 179–232.
Stapleton, R., and M. Subrahmanyam. 1984. "The Valuation of Multivariate Contingent Claims in Discrete Time Models," *Journal of Finance* 39: 207–228.
Toft, K., and B. Prucyk. 1997. "Options on Leveraged Equity: Theory and Empirical Tests," *Journal of Finance* 52: 1151–1180.
Whaley, R. 1981. "On the Valuation of American Call Options on Stocks with Known Dividends," *Journal of Financial Economics* 9: 207–211.

ABOUT THE AUTHOR

António Câmara is the Watson Family chair in Commodity and Financial Risk Management and associate professor of finance at the Spears School of Business, Oklahoma State University. He holds a *licenciatura* in business from the Universidade Catolica Portuguesa, an MBA from the Universidade Tecnica de Lisboa, and a PhD in finance from Lancaster University. He had previous appointments at the Universidade Tecnica de Lisboa, Strathclyde University, and University of Michigan, Flint. His research in derivatives has been published in many journals, including the *Journal of Finance, Journal of Business, Journal of Futures Markets*, and *Review of Derivatives Research*.

CHAPTER 28

The Pricing and Valuation of Swaps

GERALD GAY
Professor of Finance and Chairman of the Department of Finance,
Robinson College of Business, Georgia State University

ANAND VENKATESWARAN
Riesman Assistant Professor of Finance, College of Business Administration,
Northeastern University

INTRODUCTION

The size and continued growth of the global market for over-the-counter (OTC) derivative products such as swaps, forwards, and option contracts attests to their increasing and wide-ranging acceptance as essential risk management tools by financial institutions, corporations, municipalities, and government entities. Findings from a recent Bank of International Settlements survey (BIS, 2009) indicate that outstanding notional amounts of these products as of year-end 2008 surpassed \$591 trillion or \$33.9 trillion in gross market value, which represents the cost of replacing all open contracts at prevailing market prices. Of these totals, interest rate swaps alone accounted for \$419 trillion in notional amount or \$18.4 trillion in gross market value.

In this chapter we focus on this important component of the market for derivatives—swaps—and provide a primer on how they are priced and valued. While our emphasis is largely on interest rate swaps, the framework we present is applicable to a wide array of swaps including those based on currencies and commodities.[1] In addition, we provide a number of examples to illustrate such applications. The tools presented should prove useful to students of these markets having interests in trading, sales, or financial statement reporting.

A general description of a swap is that they are bilateral contracts between counterparties who agree to exchange a series of cash flows at periodic dates. The cash flows can be either fixed or floating and typically are determined by multiplying a specified notional principal amount by a referenced rate or price.

To illustrate briefly a few common types of swaps that we examine in greater detail later, a plain vanilla "fixed for floating" interest rate swap would require one party to pay a series of fixed payments based on a fixed rate of interest applied to a specified notional principal amount, while the counterparty would make variable or floating payments based on a London Interbank Offered Rate (LIBOR) interest rate applied to the same notional amount. A commodity swap involving, say, aviation jet fuel would have one party agreeing to make a series of fixed payments based on a notional amount specified in gallons multiplied by a fixed price per gallon, while the counterparty would make a series of floating payments based on an index of jet fuel prices taken from a specific geographic region. Similarly, in a currency swap, the counterparties agree to exchange two series of interest payments, each denominated in a different currency, with the added distinction that the respective principal amounts are also exchanged at maturity and possibly at origination.

Illustration 1: An End User Swap Application

To motivate our framework for pricing and valuing swaps, we first provide a hypothetical scenario involving a swap transaction. The chief financial officer (CFO) of company ABC seeks to obtain $40 million in debt financing to fund needed capital expenditures. The CFO prefers medium-term, 10-year financing at a fixed rate to provide protection against unexpected rising interest rates. The CFO faces something of a dilemma. Although the company currently enjoys an investment-grade rating of BBB, the CFO believes that its financial well-being will improve noticeably over the next couple of years to possibly an A credit rating, thus allowing it to obtain debt financing on more attractive terms. As an interim solution, the CFO considers issuing a shorter-term, floating rate note and concurrently entering a longer-term pay fixed, receive floating interest rate swap to maintain interest rate risk protection.

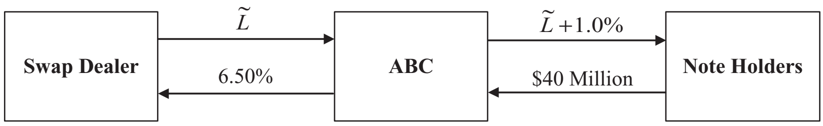
The CFO contacts a swap dealer who provides an indicative quote schedule, a portion of which is reported in Exhibit 28.1. For various swap tenors (i.e., the term of the swap), bid and ask "all-in" rates are shown.[2] We assume in the table that all quotes are presented on a semiannual, actual/365 basis versus six-month LIBOR.[3]

Consider the quote for a swap having a 10-year tenor reported to be 6.35 bid, 6.50 ask. End users paying fixed (and thus receiving floating) would make semiannual payments to the dealer based on a 6.50 percent annualized rate (ask rate) and

Exhibit 28.1 Indicative Dealer Swap Quote Schedule

	Swap Rate (%)	
Tenor (years)	**Bid**	**Ask**
1	5.75	5.80
2	5.42	5.99
3	6.05	6.12
5	6.15	6.25
7	6.25	6.35
10	6.35	6.50

Rates quoted on a semiannual, actual/365 basis versus six-month LIBOR.



ABC's Net Financing Cost:

To Creditors	$< \tilde{L}+1.0\% >$
To Swap Dealer	$< 6.50\% >$
From Swap Dealer	$+ \tilde{L}$
Net	**7.50%**

Exhibit 28.2 Illustration of Swap Transaction Fund Flows for Company ABC

using an actual/365 day count convention. End users receiving fixed would receive payments from the dealer based on the 6.35 percent bid rate. The difference in the bid and ask rates of 15 basis points represents the dealer's gross compensation for engaging in this market-making activity. The floating rate based on six-month LIBOR would be calculated using an actual/360 day count convention. Typically, the floating rate is quoted "flat" without a spread.

Upon finalization, the company issues a 3-year, \$40 million floating rate note having a quoted rate of LIBOR plus 100 basis points and concurrently enters a 10-year pay fixed, receive floating swap having \$40 million in notional amount (see Exhibit 28.2). The net effect of this set of transactions results in the company paying a synthetic fixed rate of 6.50 plus 1.00 percent or 7.50 percent for the first 3 years. For the remaining 7 years, the company would pay 7.50 percent less the reduction in its credit spread observed after 3 years when it will need to roll over its maturing note.

On a final note, the CFO would also execute an International Swaps and Derivatives Association (ISDA) Master Agreement with the swap dealer, if one is not already in place. This agreement along with its supporting schedules and addenda serves to document many of the terms and conditions governing the swap for the purpose of mitigating credit and legal risks. Once the agreement is in place, these issues need not be repeatedly negotiated upon additional transactions between the counterparties.[4]

FRAMEWORK FOR PRICING AND VALUATION

We next provide a framework for understanding swap pricing and valuation accompanied by a simple numerical example. Later we describe procedures for applying this framework when using actual market data. In swap terminology, the price of a swap differs from the value of the swap. The swap "price" refers to an interest rate, specifically, the interest rate used to determine the fixed rate payments of the swap. To begin, consider two bonds where the first bond has a fixed rate coupon while the second bond features a floating rate coupon. Values for the fixed rate bond, B^{Fix}, and the floating rate bond, B^{Flt}, are determined as shown:

$$B^{\text{Fix}} = \sum_{t=1}^{n} \frac{\overline{C}}{(1 + {}_0R_t)^t} + \frac{F}{(1 + {}_0R_n)^n} \tag{28.1}$$

$$B^{\text{Flt}} = \sum_{t=1}^{n} \frac{\widetilde{C}}{(1 + {}_0R_t)^t} + \frac{F}{(1 + {}_0R_n)^n} \tag{28.2}$$

where F = face or notional amount of each bond
$\overline{C}$ = fixed rate coupon
$\widetilde{C}_t$ = floating rate coupon associated with period t
${}_0R_t$ = rate on a zero coupon bond having a maturity t

Note that all cash flows are discounted by a unique zero coupon rate corresponding to the specific timing of the cash flow.

Next, define V to be the value of a swap. The value of a receive fixed, pay floating swap can be expressed as a portfolio consisting of a long position in a fixed rate bond and a short position in a floating rate bond. Thus, the value of the swap can be expressed as the difference in Equations 28.1 and 28.2:

$$V = B^{\text{Fix}} - B^{\text{Flt}} \tag{28.3}$$

Similarly, the value of a pay fixed, receive floating swap can be expressed as the difference in Equations 28.2 and 28.1:

$$V = B^{\text{Flt}} - B^{\text{Fix}} \tag{28.4}$$

To price the swap, we recognize two key points:

1. At its inception, the value of a fairly priced swap is zero.
2. The value of a floating rate bond at either issuance or upon any reset date is its par or face amount. For discussion purposes, we assume the par amount equals \$1.

Thus, using either Equation 28.3 or 28.4, we have:

$$\begin{aligned} V = B^{\text{Fix}} - B^{\text{Flt}} &= \$0; \\ B^{\text{Fix}} - \$1 &= \$0; \text{ thus,} \\ B^{\text{Fix}} &= \$1 \end{aligned} \tag{28.5}$$

Expression 28.5 provides the key insight into pricing a swap. The "price" of a swap (sometimes referred to as the par value swap rate) will be the coupon rate that makes the fixed rate bond have a value equal to that of the floating rate bond and thus causes the initial swap value to equal zero.

Illustration 2: A Simple Example

Consider an example wherein we seek to price a one-year (360-day) swap having semiannual payments at 180-day intervals and a \$1 notional amount. As we discuss later in greater detail, we use LIBOR interest rates to discount cash flows and follow money market conventions by using an actual/360-day count convention. Since cash flows occur at two dates (180 and 360 days), we assume that ${}_0L_{180} =$ 6.00 percent and ${}_0L_{360} = 7.00$ percent. To price the swap, we simply solve for the fixed coupon rate $\overline{C}$ that causes B^{Fix} to sell at par, according to the next expression:

$$B^{Fix} = \frac{\overline{C} \times \dfrac{180}{360}}{\left(1 + 0.06 \times \dfrac{180}{360}\right)} + \frac{\overline{C} \times \dfrac{180}{360} + 1}{\left(1 + 0.07 \times \dfrac{360}{360}\right)} = 1$$

Solving for $\overline{C}$, we find its value to equal 0.0686. The price of the swap is thus 0.0686 or 6.86 percent.[5] The fixed rate payer will therefore make a fixed payment each period of \$0.0343 per \$1 of notional value.

For the floating rate payer, standard convention is that floating payments are based on LIBOR rates observed at the start of the period rather than in arrears. Thus, the initial floating payment will be 6.00% × 180/360 × \$1 or \$0.0300. The amount of the second floating payment will be determined at the next settlement date when a new value of ${}_0L_{180}$ is observed.

Next, we consider how to value a swap at times following origination. Both the passage of time and changing market interest rates can cause the swap to take on positive or negative value. Assume in this same example that 10 days elapse during which market interest rates appeared to have risen. The two remaining settlement dates are now in 170 and 350 days, respectively; thus we require zero-coupon discount rates corresponding to 170 and 350 days. Assume that ${}_0L_{170} =$ 6.50 percent and ${}_0L_{350} = 7.50$ percent. To solve for the new value of the swap we first estimate the new values of the two component bonds, B^{Fix} and B^{Flt}.

Consider the floating rate bond. Since only 10 days have elapsed, the \$.0300 coupon to be paid at the next settlement date is still in effect. Also, upon the next settlement date, a new floating rate will be selected such that the remaining value of the floating rate bond is restored to par. The value of B^{Flt} is determined as shown next:

$$B^{Flt} = \frac{0.0300 + 1}{\left(1 + 0.0650 \times \dfrac{170}{360}\right)} = 0.99933$$

For the fixed rate bond, there are two remaining coupons of \$.0343 along with the repayment of the principal amount. The value of B^{Fix} is determined as:

$$B^{Fix} = \frac{0.0343}{\left(1 + 0.0650 \times \dfrac{170}{360}\right)} + \frac{0.0343 + 1}{\left(1 + 0.0750 \times \dfrac{350}{360}\right)} = 0.99729$$

Thus, the value of the swap, assuming pay fixed, receive floating, is equal to:

$$V = B^{Flt} - B^{Fix} = 0.99933 - 0.99729 = \$0.00204$$

For the same swap having instead a notional amount of \$50 million, the value of the swap would be 50,000,000 times \$0.00204 or +\$102,000. This is the value to the fixed rate payer. To the fixed rate receiver, a receive fixed, pay floating swap would have a value of –\$102,000. The change in value of \$102,000 also represents the profit/loss on the swap over the 10-day interval due to the change in interest rates.[6]

STEPS FOR SWAP PRICING

In this section we extend the preceding procedures to apply to market data and to reflect industry conventions. But first we consider credit concerns. Common practice in swap pricing is to use LIBOR-based rates to discount cash flows and to determine floating payments. LIBOR reflects a rate at which high-quality borrowers, typically A- and AA-rated commercial banks, can obtain financing in capital markets. While most swap counterparties are of investment-grade quality, significant differentials in their credit quality can exist. Rather than make adjustments to the swap price or rate, these differences typically are accounted for through nonprice means as specified in master agreements.

As mentioned earlier, most swaps will be executed under a master agreement, which contains a number of provisions aimed at mitigating credit and legal risk. Master agreements typically specify terms that allow counterparties to engage in payment netting and to conduct both (1) up-front risk assessment through the provision of documents concerning credit risk as well as through representations concerning enforceability; and (2) ongoing risk assessment through the periodic provision of documents, maintenance of covenants, use of collateral, and mark-to-market margining.

Next we discuss these steps for pricing and valuing swaps:

- Obtain market inputs.
- Make convexity adjustments to implied futures rates.
- Build the zero curve.
- Identify relevant swap features.
- Price/value the swap.

Obtain Market Inputs

Because we will be computing the present value of swap cash flows, for each payment date, we require a discount rate of a corresponding maturity. These rates will be drawn from a LIBOR-based, zero-coupon discount curve, which henceforth we refer to as the zero curve. Since most of the rates comprising this curve are not readily available, they must be estimated using market data. For rates at the short end of the zero curve, LIBOR spot rates are directly observable and easy to obtain. For the middle and longer-term portions of the zero curve, the rates can be derived from forward rate agreements (FRAs) or, as illustrated next, Eurodollar futures.

Eurodollar futures, traded on the Chicago Mercantile Exchange, are among the world's most liquid and heavily traded futures due in large part to their central role in the pricing and hedging of interest rate swaps. Eurodollar futures are in essence cash-settled futures on three-month LIBOR-based deposit rates. Quotes on Eurodollar futures prices are readily available for maturities extending out 10 years. Prices are quoted such that to find the associated implied futures rate, one must first subtract the observed futures price from 100.00. Consider a Eurodollar futures quoted at 94.00. To find its associated implied futures rate, we subtract this price from 100.00 and obtain a rate of 6.00 percent.[7]

For use in the calculations to follow, we present in Exhibit 28.3 a set of sample spot LIBOR rates and Eurodollar futures prices observed on December 18, 2007. One-, three-, and six-month LIBOR spot rates are given. In addition, we report Eurodollar futures prices and corresponding implied futures rates, beginning with the June 2008 contract and extending to the September 2011 contract. In practice, the termination of trading of a Eurodollar futures contract is two business days prior to the third Wednesday of the contract month; we abstract somewhat and assume for purposes of our analysis that this date will fall on the eighteenth day of each contract month. Also, given that the three-month implied futures rate underlying the September 2011 contract extends from September 18, 2011, to December 18, 2011, the market data reported in Exhibit 28.3 permits us to price and value swaps extending out four years. To price swaps having even longer tenor, we would simply add futures having more distant maturities.

Make Convexity Adjustments to Implied Futures Rates

We use implied Eurodollar futures rates to proxy for the forward rates that are used to construct the zero curve. However, due to the daily resettlement feature common to futures markets (and not to forward markets), the implied futures rate is likely to be an upward-biased estimate of the desired forward rate. To see why this is the case, consider the next illustration.

Assume that a Eurodollar futures contract is trading at 95.00 (implied futures rate of 5 percent). For each contract, there will be a long and short position. If the futures price increases 100 basis points to 96.00 (implied futures rate of 4 percent), then, following daily resettlement, the long position will show a credit or profit of $2,500 (100 basis points times $25 per basis point). These funds are then invested elsewhere to earn 4 percent. The account of the short position will show a debit or loss of $2,500, which is financed at 4 percent. Consider instead the result of a 100 basis point decline in price to 94.00 (implied futures rate of 6 percent). In this case, the long position incurs a loss of $2,500, which is financed at 6 percent. The short position earns a $2,500 profit, which is invested at 6 percent.

Assuming that positive and negative price changes are equally likely, this situation clearly favors the short position. This is because short (long) profits and thus invests when rates are relatively higher (lower), and loses and thus borrows when rates are relatively lower (higher). Longs recognize this and thus require compensation to induce them to trade. Thus, shorts will agree to a lower equilibrium price than they would otherwise in the absence of daily resettlement. As a result of the lowered futures price, the implied futures rate becomes an upward-biased proxy for the desired forward rate. This bias will become more severe (1) the longer the

Exhibit 28.3 Market Data for the Zero Curve Construction (for December 18, 2007)

(a) LIBOR Spot Rate Information

Term	Maturity	Days	Rate (%)
1 month	18-Jan-08	31	4.9488
3 month	18-Mar-08	91	4.9263
6 month	18-Jun-08	183	4.8250

(b) Eurodollar Futures Information

Contract	Price	Start Date	Days (T1)	End Date	Days (T2)	Days (T1 – T2)	Futures Rate (%)	Convexity Adjustment	Forward Rate (%)
Jun 2008	96.070	18-Jun-08	183	18-Sep-08	275	92	3.9300	0.0019	3.9281
Sep 2008	96.310	18-Sep-08	275	18-Dec-08	366	91	3.6900	0.0038	3.6862
Dec 2008	96.410	18-Dec-08	366	18-Mar-09	456	90	3.5900	0.0063	3.5837
Mar 2009	96.405	18-Mar-09	456	18-Jun-09	548	92	3.5950	0.0094	3.5856
Jun 2009	96.295	18-Jun-09	548	18-Sep-09	640	92	3.7050	0.0132	3.6918
Sep 2009	96.155	18-Sep-09	640	18-Dec-09	731	91	3.8450	0.0176	3.8274
Dec 2009	96.025	18-Dec-09	731	18-Mar-10	821	90	3.9750	0.0225	3.9525
Mar 2010	95.905	18-Mar-10	821	18-Jun-10	913	92	4.0950	0.0281	4.0669
Jun 2010	95.770	18-Jun-10	913	18-Sep-10	1005	92	4.2300	0.0344	4.1956
Sep 2010	95.660	18-Sep-10	1005	18-Dec-10	1096	91	4.3400	0.0413	4.2987
Dec 2010	95.560	18-Dec-10	1096	18-Mar-11	1186	90	4.4400	0.0488	4.3912
Mar 2011	95.480	18-Mar-11	1186	18-Jun-11	1278	92	4.5200	0.0569	4.4631
Jun 2011	95.395	18-Jun-11	1278	18-Sep-11	1370	92	4.6050	0.0657	4.5393
Sep 2011	95.310	18-Sep-11	1370	18-Dec-11	1461	91	4.6900	0.0751	4.6149

time to the futures maturity date because of the greater number of daily resettlements; and (2) the longer the duration or price sensitivity of the asset underlying the futures contract.

To correct for the bias, a convexity adjustment is applied to the futures rate. A reasonable and simple approximation for the convexity adjustment is given by:[8]

$$\text{Forward rate} = \text{Futures rate} - (.5) \times \sigma^2 \times \text{T1} \times \text{T2} \tag{28.6}$$

where σ = annualized standard deviation of the change in the short-term interest rate
T1 = time (expressed in years) to the futures maturity date
T2 = time to the maturity date of the implied futures rate

We report in Exhibit 28.3 the results of convexity adjustments to all the implied futures rates assuming a value of σ of 1 percent. To see the potential magnitude of the convexity adjustment, consider first the September 2008 contract reported at a price of 96.31 (implied futures rate of 3.69 percent). This contract matures in 275 days while the implied futures rate extends out another 91 days to day 366. Thus,

$$\begin{aligned}\text{Forward rate} &= .036900 - .5 \times .01^2 \times 275/365 \times 366/365 \\ &= .036900 - .000038 = .036862\end{aligned}$$

The convexity adjustment in this instance is quite small at .38 basis points. Consider, however, the convexity adjustment for a longer dated futures such as the September 2011 contract quoted at 95.31 (implied futures rate of 4.69). Thus,

$$\begin{aligned}\text{Forward rate} &= .046900 - .5 \times .01^2 \times 1370/365 \times 1461/365 \\ &= .046900 - .000751 = .046149\end{aligned}$$

In this instance, the convexity adjustment was a more significant 7.51 basis points. For a Eurodollar futures maturing in 8 years (not shown in Exhibit 28.3) the adjustment would grow to $.5 \times .01^2 \times 8 \times 8.25 = .0033$, or 33 basis points. Thus, it is important to make the convexity adjustment to implied futures rates when pricing swaps, especially swaps of longer maturities.

Build the Zero Curve

To construct the zero curve, we use a bootstrapping technique in which short-term LIBOR spot rates are combined with forward rates to produce LIBOR-based, discount rates of longer maturities. Given that Eurodollar futures contract maturities extend out 10 years, it is possible to use this technique to produce rates of similar maturity.

Consider the spot and forward rates reported in Exhibit 28.3. We see that the six-month LIBOR spot rate (${}_0\text{L}_{183}$) is 4.8250 percent, and extends out 183 days to June 18, 2008. The forward rate associated with the June 2008 futures contract is 3.9281 percent (${}_{183}\ell_{92}$) and commences on June 18 and extends 92 days to

September 18.[9] Thus, we can compute a 275-day LIBOR-based discount rate by linking the two rates:

$$(1 + {}_0L_{275} \times 275/360) = (1 + {}_0L_{183} \times 183/360) \times (1 + {}_{183}\ell_{92} \times 92/360)$$
$$(1 + {}_0L_{275} \times 275/360) = (1 + 0.048250 \times 183/360) \times (1 + .039281 \times 92/360)$$
$${}_0L_{275} = .045572, \text{ or } 4.5572\%$$

Using this 275-day rate, we can similarly calculate a 366-day LIBOR rate by linking it to the 91-day forward rate associated with the September 2008 futures contract:

$$(1 + {}_0L_{366} \times 366/360) = (1 + {}_0L_{275} \times 275/360) \times (1 + {}_{275}\ell_{91} \times 91/360)$$
$$(1 + {}_0L_{366} \times 366/360) = (1 + 0.045572 \times 275/360) \times (1 + .036862 \times 91/360)$$
$${}_0L_{366} = .043725, \text{ or } 4.3725 \text{ percent}$$

We repeat this process in an iterative manner to produce a set of LIBOR-based discount rates extending out essentially four years and at three-month intervals. Given the data available in the table, the longest term discount rate (${}_0L_{1461}$) is computed as:

$$(1 + {}_0L_{1461} \times 1461/360) = (1 + {}_0L_{1370} \times 1370/360) \times (1 + {}_{1370}\ell_{91} \times 91/360)$$
$$(1 + {}_0L_{1461} \times 1461/360) = (1 + 0.044361 \times 1370/360) \times (1 + .046149 \times 91/360)$$
$${}_0L_{1461} = .044957, \text{ or } 4.4957 \text{ percent}$$

Following these procedures, we report in Exhibit 28.4 our set of discount rates representing the zero curve.

Exhibit 28.4 Zero Curve Rate Information (for December 18, 2007)

Maturity	Term (Days)	Rate (%)
18-Jan-08	31	4.9488
18-Mar-08	91	4.9263
18-Jun-08	183	4.8250
18-Sep-08	275	4.5572
18-Dec-08	366	4.3725
18-Mar-09	456	4.2483
18-Jun-09	548	4.1694
18-Sep-09	640	4.1345
18-Dec-09	731	4.1313
18-Mar-10	821	4.1480
18-Jun-10	913	4.1786
18-Sep-10	1005	4.2209
18-Dec-10	1096	4.2694
18-Mar-11	1186	4.3219
18-Jun-11	1278	4.3778
18-Sep-11	1370	4.4361
18-Dec-11	1461	4.4957

Before concluding this discussion, we note two situations where it will be necessary to interpolate between rates.

1. We assumed a start date (December 18, 2007) in which the six-month LIBOR spot rate extended exactly to a futures maturity date (June 18, 2008), thus allowing us to link immediately to a forward rate. In practice, this will usually not be the case, and one will need to interpolate between two spot rates surrounding a futures maturity date in order to link to the first forward rate. To illustrate, assume the start date is February 1, 2008. The three-month LIBOR rate will extend to 90 days to May 1 while the six-month rate will extend 182 days to August 1. What is required, however, is a spot rate extending 138 days to June 18. Assume that the three- and six-month LIBOR rates are 5.00 and 6.00 percent, respectively. Through linear interpolation, we estimate the 138-day rate as $5.00 + (6.00 - 5.00) \times (138 - 90)/(182 - 90) = 5.52\%$. If one had access to four- and five-month LIBOR spot rates extending to June 1 and July 1, respectively, a more accurate interpolation is then possible.
2. The computed rates comprising the zero curve had maturities corresponding to the futures maturity dates. More than likely, these dates will not be the same as the swap payment dates. Thus, one will need to interpolate between the computed zero curve rates to produce additional discount rates that correspond to the swap payment dates.[10]

Identify Relevant Swap Features

Key information required to price a swap includes its tenor, settlement frequency, payment dates, and day count conventions. To value an existing swap, one also needs to know the notional value of the swap, the swap rate, the floating rate in effect for the current payment period, and whether the swap is to be valued from the either the fixed rate payer or receiver perspective.

Swap settlement frequencies of quarterly and semiannual are most common, while monthly and annual frequencies are also observed. The day count convention addresses how days in the payment period will be counted and how many days are assumed to be in a year. For purposes of calculating the fixed rate interest payment, the three most common day count conventions used in U.S. dollar swaps are Actual/360, Actual/365, and 30/360. The numerator in each refers to the number of days assumed to be in the payment period while the denominator specifies the number of days assumed to be in a year. "Actual," as its name suggests, refers to the actual number of days in the period. The 30/360 convention, sometimes referred to as bond basis, assumes that every month contains 30 days regardless of the actual number of days in the month. For purposes of calculating the floating interest payment, the Actual/360 convention typically is used since it will be based on a LIBOR interest rate.

Price/Value the Swap

Following the completion of the preceding steps, one is prepared to price a swap to be originated or to value an existing swap. Using the information presented in Exhibit 28.4, we illustrate the pricing of a swap followed by the valuation of an existing swap.

Exhibit 28.5 Swap Rates (%) (Rates Based on Market Data from December 18, 2007)

	Day Count Convention		
Tenor (Years)	**Actual/360**	**Actual/365**	**30/360**
1	4.3304	4.3906	4.4026
2	4.0159	4.0717	4.0773
3	4.0582	4.1145	4.1183
4	4.1696	4.2275	4.2304

Illustration 3: Pricing an Interest Rate Swap

Consider a swap having a three-year tenor, semiannual payments, and an Actual/365 day count convention. The payment dates thus fall on June 18 and December 18 of each year. To price the swap, we find the value $\overline{C}$ in Equation 28.7 that causes a bond with a face value of \$1 to sell for par.

$$\$1 = \frac{\overline{C} \times 183/365}{\left(1 + 0.048250 \times \frac{183}{360}\right)} + \frac{\overline{C} \times 183/365}{\left(1 + 0.043725 \times \frac{366}{360}\right)} + \frac{\overline{C} \times 182/365}{\left(1 + 0.041694 \times \frac{548}{360}\right)} + \frac{\overline{C} \times 183/365}{\left(1 + 0.041313 \times \frac{731}{360}\right)} + \frac{\overline{C} \times 182/365}{\left(1 + 0.041786 \times \frac{913}{360}\right)} + \frac{\overline{C} \times 183/365 + 1}{\left(1 + 0.042694 \times \frac{1096}{360}\right)} \tag{28.7}$$

Solving for $\overline{C}$, the swap rate is 4.1145 percent. Note that for each coupon payment, we apply an actual/365 day count convention and a unique discount rate. Exhibit 28.5 reports this and other swap rates for tenors ranging from one to four years under each of the various day count conventions.

Illustration 4: Valuing an Existing Interest Rate Swap

Consider the valuation of an existing swap again using the rate information presented in Exhibit 28.4. Assume that a corporation entered a pay fixed/receive floating swap a few years ago at a swap rate of 2.00 percent (Actual/365) when rates were significantly lower than currently. Assume also that the swap has a notional amount of \$40 million, makes semiannual payments on March 18 and September 18 of each year, and is set to mature on March 18, 2009. Hence, there are three payment dates remaining. The six-month LIBOR rate in effect for the current period is 5.4200 percent, which was last reset on September 18, 2007.

To value the swap, we value a portfolio of two bonds in which we assume the swap holder is long a floating rate bond and short a fixed rate bond. As discussed earlier, the value of the floating rate bond will be restored to its par amount of \$1 upon the next reset date. Thus, to value the floating rate bond, we simply find the present value of the sum of \$1 plus the next semiannual floating rate payment of $.0542 \times 182/360 = .02740$ to be received in 91 days on March 18, 2008. Thus,

$$B^{Flt} = \frac{0.02740 + 1}{\left(1 + 0.049263 \times \frac{91}{360}\right)} = 1.014764$$

For the fixed rate bond, we find the present value of the three remaining semiannual payments plus the $1 face value. Thus,

$$B^{Fix} = \frac{.0200 \times 182/365}{\left(1 + 0.049263 \times \frac{91}{360}\right)} + \frac{0.0200 \times 184/365}{\left(1 + 0.045572 \times \frac{275}{360}\right)} + \frac{0.0200 \times 181/365 + 1}{\left(1 + 0.042483 \times \frac{456}{360}\right)}$$
$$= 0.977940$$

The value of the swap is thus:

$$V = B^{Flt} - B^{Fix} = [1.014764 - 0.977940] \times \$40{,}000{,}000$$
$$= 0.036824 \times \$40{,}000{,}000 = \$1{,}472{,}960$$

The valuation of this swap could serve a number of purposes. For example, for purposes of financial statement reporting, the corporation would record the swap as an asset valued at $1,472,960, while its counterparty would record the same value as a liability. Also, if the two parties agreed to terminate the swap, the corporation would receive a payment of $1,472,960 from the counterparty.

OTHER SWAPS

The framework just discussed can be extended easily to swaps other than interest rate swaps. We next discuss such applications to currency and commodity swaps.

Currency Swaps

The interest rate swaps just considered can be thought of as single-currency swaps, as all cash flows were denominated in the same currency, for example, U.S. dollars. Importantly, the procedures we reviewed are equally applicable when working with single-currency interest rates swaps denominated in currencies other than the U.S. dollar.

Currency swaps are in essence interest rate swaps wherein the two series of cash flows exchanged between counterparties are denominated in typically two different currencies. The interest payments can be in a fixed-for-floating, fixed-for-fixed, or floating-for-floating format.[11] Floating rates typically are expressed in terms of a LIBOR rate based on a specified currency.[12] An added distinction is that the principal amounts in a currency swap are not merely notional but rather typically are actually exchanged at maturity and also may be exchanged when the swap is originated.

The procedures for pricing and valuing currency swaps become intuitive when considered in the context of our earlier discussion of interest rate swaps. Regardless of the currency denomination of any single-currency swap, these three rules should hold.

1. The swap can be viewed as a portfolio of two bonds, a fixed rate bond and a floating rate bond, wherein one of the bonds is held long and the other is held short.
2. The initial value of the swap is zero since the value of the fixed rate bond equals that of the floating rate bond.

3. The fixed swap rate, when set correctly, is the coupon rate that makes the fixed rate bond sell at par. Thus, from a value perspective, at origination, one would be indifferent between holding long or short either bond comprising the swap since both bonds have identical value equal to their par or notional principal amounts.

In a currency swap, one party is in essence long one of the two bonds comprising a single-currency interest rate swap of one currency and is short one of the two bonds comprising a single-currency swap of a second currency. The counterparty's position is opposite. To make each party indifferent as to the two bonds they hold representing the two legs of the currency swap, the bonds' par or notional principal amounts are set to reflect the current spot exchange rate. This in turn will lead to an initial value of the swap equal to zero. Thus, at the swap's origination, the two principal amounts are such that:

$$B_0{}^{Dom} = B_0{}^{For} \times S_0 \tag{28.8}$$

where $B_0{}^{Dom}$ = initial value or principal amount of the bond having cash flows (fixed or floating) expressed in the domestic currency
$B_0{}^{For}$ = initial value or principal amount of the bond having cash flows (fixed or floating) expressed in the foreign currency
S_0 = current spot exchange rate (Dom/For)

Again, note that at the origination of the swap, the value of each bond will equal their respective par or principal amounts.

Illustration 5: Pricing and Valuing a Currency Swap

To illustrate the pricing of a currency swap, consider the following market information from Wednesday, December 26, 2007, involving the U.S. dollar and Swiss franc. For five-year U.S. dollar swaps having semiannual payments, the swap rate was 4.38 percent, the five-year Swiss franc swap rate was 3.05 percent, and the spot exchange rate was .8687 (US/SF). Now consider a currency swap in which the U.S. dollar leg has a principal amount of $20 million. At the current exchange rate, this is equivalent to approximately SF23.02 million. A number of currency swap structures are then possible: fixed for fixed, fixed for floating, or floating for floating. If properly priced and configured, each swap would entail the next four cash flows between the counterparties, Party One and Party Two:

1. At origination, the counterparties may exchange the two principal amounts; if so, Party One would send $20 million to Party Two while receiving SF23.02 million from Party Two.
2. Party One would receive from Party Two U.S. dollar interest payments based on a 4.38 percent fixed rate over the life of the swap or, alternatively, the U.S. dollar LIBOR floating rate. Either rate would be applied to a principal amount of $20 million.
3. Party Two would receive from Party One Swiss franc interest payments based on a 3.05 percent fixed rate over the life of the swap or, alternatively,

the SF LIBOR floating rate. Either rate would be applied to a principal amount of SF23.02 million.

4. At maturity, and regardless of whether there was an initial exchange, the counterparties would exchange the original principal amounts. Party One would receive \$20 million from Party Two and send SF23.02 million to Party Two.

To value the swap at times subsequent to origination, note that a currency swap can take on positive or negative value depending on changes in the spot exchange rate and in the interest rates associated with each currency. Thus, one simply finds the present value of each respective set of cash flows using the revised zero curve for that currency and accounts for the prevailing exchange rate. The value of a receive domestic/pay foreign swap at any time t (V_t) expressed in terms of the domestic currency is thus equal to:

$$V_t = B_t^{Dom} - \left\{B_t^{For} \times S_t\right\} \tag{28.9}$$

while the value of a pay domestic/receive foreign swap equals:

$$V_t = \left\{B_t^{For} \times S_t\right\} - B_t^{Dom} \tag{28.10}$$

Commodity Swaps

Finally, we consider the pricing and valuation of commodity swaps. Note upon substituting Equations 28.1 and 28.2 for B^{Fix} and B^{Flt}, respectively, into Equation 28.3 and simplifying, the value of a swap can also be written as:

$$V = \sum_{t=1}^{n} \frac{\overline{C} - \widetilde{C}}{(1 + {}_0R_t)^t} \tag{28.11}$$

This equation is appealing as it expresses the swap's value in terms of the portfolio value of a series of n forward contracts. Equation 28.11 is especially useful for pricing a commodity swap in that one simply solves for the fixed price of the commodity $\overline{C}$ that makes the overall value of the swap equal zero. That is, one solves for the value of $\overline{C}$ that does not make the value of each of the n forward contracts equal zero but rather the sum of the value of all n forward contracts.

Before applying Equation 28.11, one must first build a zero curve as previously shown. In addition, one must also obtain a set of forward prices for the commodity corresponding to each payment date of the swap. Typically, these can be taken from the commodity's forward curve, or futures prices can be used.

Illustration 6: Pricing a Commodity Swap

We again assume that the current date is December 18, 2007, which allows us to make use of the zero curve information presented in Exhibit 28.5. We consider an end user of crude oil that wishes to fix its supply costs over the next two years.

The end user approaches a commodity swap dealer and is presented with the next term sheet.

Commodity:	**CrudeoOil (West Texas Intermediate)**
Notional Amount:	100,000 barrels
Agreed Fix Price:	\$87.59/barrel
Agreed Oil Price Index:	"Oil-WTI-Platt's Oilgram"
Term:	2 years
Settlement Basis:	Cash settlement, Semiannual
Payment Dates:	June 18 and December 18

To determine whether the quoted \$87.59 swap price reflects current market conditions, the end user observes that current New York Mercantile Exchange (NYMEX) West Texas Intermediate crude oil futures prices are in their typical backwardation pattern. Futures prices corresponding to the next four semiannual dates are: \$89.50 (June 2008), \$88.00 (December 2008), \$86.75 (June 2009), and \$86.00 (December 2009). Applying Equation 28.11, we solve for the value of $\overline{C}$ in the next expression, which causes the overall value of the swap to equal zero.

$$0 = \frac{\overline{C} - \$89.50}{\left(1 + 0.048250 \times \frac{183}{360}\right)} + \frac{\overline{C} - \$88.00}{\left(1 + 0.043725 \times \frac{366}{360}\right)} + \frac{\overline{C} - \$86.75}{\left(1 + 0.041694 \times \frac{548}{360}\right)} + \frac{\overline{C} - \$86.00}{\left(1 + 0.041313 \times \frac{731}{360}\right)}$$

Solving, we find that $\overline{C}$ equals \$87.59 and conclude that the swap is priced correctly.

As we mentioned at the outset, the market for OTC derivatives continues to grow rapidly, reflecting their value and acceptance as important risk management tools. Focusing on the largest segment of this market, swaps, we have attempted to present a simple framework for facilitating an understanding of their pricing and valuation, relying mainly on time value of money concepts and noting a few market conventions.

ENDNOTES

1. For a review and analysis of other popular swap structures including credit default swaps, equity swaps, and total return swaps, see, for example, Bomfim (2005), Chance and Rich (1998), Chance and Brooks (2007), Choudhry (2004), and Kolb and Overdahl (2007).
2. Alternatively, quotes may be presented in terms of a swap spread, an amount to be added to the yield of a Treasury instrument having a comparable tenor. The swap spread should not be confused with the bid-ask spread of the swap quote.
3. Later we discuss and consider other common day count conventions.
4. For additional discussion regarding the role of the ISDA Master Agreement, see Gay and Medero (1996).

5. In practice, 6.86 percent would be the mid-rate around which a swap dealer would establish a bid-ask quote.
6. For financial reporting purposes, according to Financial Accounting Statement No. 133, if day 10 corresponded to an end of period reporting date, an end user would record the pay fixed, receiving floating swap as a \$102,000 asset, while the counterparty would record the swap as a \$102,000 liability. The change in value of the swap since origination would also have income statement implications depending on the swap's intended purpose.
7. The standard contract size for Eurodollar futures is \$1 million notional. Thus, a 1 basis point change in price corresponds to a $\$1{,}000{,}000 \times .0001 \times 90/360 = \$25$ change in contract value.
8. See Hull (2008, pp. 136–138) as well as Ron (2000), who provides a more elaborate estimation for the convexity adjustment. Also, Gupta and Subrahmanyam (2000) conduct an empirical investigation into the extent that the market has correctly incorporated over time the convexity adjustment into observed swap rates.
9. In our notation we use an upper case "L" to denote a LIBOR-based spot rate and a lower case "ℓ" to denote a LIBOR-based forward rate.
10. For additional discussion and detail regarding curve construction, see Overdahl, Schachter, and Lang (1997).
11. See Kolb and Overdahl (2007) for additional analysis of various currency swap structures.
12. In addition to the U.S. dollar, LIBOR rates are specified in the several other currency denominations, including the Australian dollar, British pound sterling, Canadian dollar, Danish krone, euro, Japanese yen, New Zealand dollar, Swedish krona, and Swiss franc.

REFERENCES

Bank of International Settlements. 2009. "OTC Derivative Market Activity in the Second Half of 2008."

Bomfim, A. 2005. *Understanding Credit Derivatives and Related Instruments*. San Diego, CA: Elsevier Academic Press.

Chance, D., and R. Brooks. 2007. *An Introduction to Derivatives and Risk Management*, 7th ed. Mason, OH: Thomson South-Western.

Chance, D., and D. Rich. 1998. "The Pricing of Equity Swaps and Swaptions," *Journal of Derivatives* 5 (Summer): 19–31.

Choudhry, M. 2004. *An Introduction to Credit Derivatives*. Oxford: Elsevier Butterworth-Heinemann.

Gay, G., and J. Medero. 1996. "The Economics of Derivatives Documentation: Private Contracting as a Substitute for Government Regulation," *Journal of Derivatives* 3: 78–89.

Gupta, A., and M. Subrahmanyam. 2000. "An Empirical Examination of the Convexity Bias in the Pricing of Interest Rate Swaps," *Journal of Financial Economics* 55: 239–279.

Hull, J. 2008. *Fundamentals of Futures and Options Markets*, 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Kolb, R., and J. Overdahl. 2007. *Futures, Options, and Swaps*, 5th ed. Malden, MA: Blackwell Publishing.

Overdahl, J., B. Schachter, and I. Lang. 1997. "The Mechanics of Zero-Coupon Yield Curve Construction," in A. Cornyn, R. Klein, and J. Lederman, eds., *Controlling & Managing Interest-Rate Risk*. New York: New York Institute of Finance.

Ron, U. 2000. "A Practical Guide to Swap Curve Construction," Working Paper 2000-17, Financial Markets Department, Bank of Canada.

ABOUT THE AUTHORS

Gerald Gay is professor of finance and chairman of the Department of Finance in the Robinson College of Business at Georgia State University. He earned BSE and MSE degrees in industrial systems engineering from the University of South Florida and a PhD in finance from the University of Florida. His research and professional interests center on financial risk management and the valuation and use of derivative instruments. He has authored over 50 articles in leading academic journals. He currently services as the coeditor of the *Journal of Financial Research* and is an associate editor of the *Journal of Futures Markets*. Previously he served as the chief economist of the U.S. Commodity Futures Trading Commission in Washington, DC, and as the director of its Division of Economic Analysis. He also serves as an advisor and provider of executive training to corporations and the legal profession.

Anand Venkateswaran is a Riesman Assistant Professor of Finance in the College of Business Administration at Northeastern University. He earned his undergraduate degree in chemical engineering from the University of Pune in India and a PhD in finance from Georgia State University in Atlanta. Prior to his doctoral studies, he worked as an investment banker for five years. Dr. Venkateswaran's primary research interests are in the area of corporate finance with a focus on managerial incentives. His primary teaching interests are in corporate finance, international finance, and risk management. He has published articles in the *Journal of Finance* and the *Journal of Financial Research*.

PART V

Advanced Pricing Techniques

While all pricing of financial derivatives exploits the no-arbitrage principle, and while many derivatives can be priced with closed-form models in the spirit of the original Black-Scholes model, the pricing of other derivatives requires different techniques, which Part Five explores. Cara M. Marshall shows how to price various derivatives using the technique of Monte Carlo analysis in Chapter 29, "Monte Carlo Techniques in Pricing and Using Derivatives." In essence, Monte Carlo analysis involves applying appropriate rules to create many potential outcomes. Given a large sample of outcomes, one can determine the likely payoffs on a financial derivative by applying probability concepts, and given an estimate of the payoffs, one can estimate the current price of the financial derivative.

Lattice or finite difference models provide another technique for pricing financial derivatives, and this method has proven extremely powerful in pricing almost all kinds of financial derivatives, as Craig Pirrong explains in Chapter 30, "Valuing Derivatives Using Finite Difference Methods." Essentially, finite difference methods proceed by breaking a span of time into many discrete time intervals and computing how the value of a financial derivative would evolve backward in time from the known payoffs at expiration to the present. For example, if we can specify a reasonable distribution of stock prices that will prevail at the expiration of an option, we know what the payoff on the option will be, contingent on those stock prices. The finite difference approach steps back a discrete time interval (back from the expiration date and closer to the present) to find the value of the option at that time. It repeats this process until the current time is reached and the value of the option at the present moment is computed. Pirrong explains in detail how this process works and shows how these finite difference methods can be applied in more complicated pricing situations.

In Chapter 31, "Stochastic Processes and Models," George Chalamandaris and A. G. Malliaris introduce the reader to definitions and key properties of stochastic processes that are important in finance. The authors start their analysis by focusing on the stochastic process known as Brownian motion, which describes the idea of a continuous random walk, and proceed to Ito processes, which incorporate both trend and volatility. In their exposition, Chalamandaris and Malliaris emphasize how to apply stochastic processes in financial modeling. They show why ordinary calculus cannot tackle the problems that arise in continuous time financial economics because of the presence of randomness, and they offer a brief presentation of the main concepts of stochastic calculus by reviewing the Ito integral and the Ito formula.

In the standard Black-Scholes-Merton option pricing model, the value of an option depends on the price of the underlying stock, the volatility of that stock, the exercise price, the time to expiration, the interest rate, and the continuous dividend rate. Understanding how the option price responds to changes in these variables is an essential part of understanding option pricing, as R. Brian Balyeat explains in Chapter 32, "Measuring and Hedging Option Price Sensitivities." Balyeat shows how to compute these sensitivities for each of the input parameters and illustrates how the option price responds to each. He also shows how to compute and analyze these sensitivities for portfolios of options, which leads naturally to a discussion of the importance of these sensitivities in portfolio management.

CHAPTER 29

Monte Carlo Techniques in Pricing and Using Derivatives

CARA M. MARSHALL
Lecturer of Economics/Finance at Queens College, City University of New York

INTRODUCTION

Of all financial products, the most difficult to price are the derivatives. Of course, some derivatives are far more difficult to price than others. The models used to price derivatives can be divided into three broad categories: analytical models, numeric models, and simulation models.[1] The latter refers to Monte Carlo simulation, named after a famous casino in Monaco on the French Riviera. Monte Carlo simulation was first used to estimate the probability of winning a game of pure chance. The technique works by replicating the outcomes of a stochastic process through the careful use of random numbers. As the number of replications increases, the resulting range of approximation, represented by the average, narrows, converging to the analytically correct solution. Since Monte Carlo simulation generally requires a significant number of repetitions of an algorithm (a massive number of calculations), its practical application requires a computer with a fast CPU. Monte Carlo simulation can be used in all sorts of business applications whenever there is a source of uncertainty (such as future stock prices, interest rates, exchange rates, commodity prices, etc.). To illustrate the basic concepts, we focus on pricing options, which are generally the most difficult types of derivatives to value.

The classic example of an analytical model used to price options is the Black-Scholes model published by Fischer Black and Myron Scholes in 1973. The classic example of a numeric model to price options is the binomial option pricing model published by John Cox, Stephen Ross, and Mark Rubinstein in 1979.

Analytical models are the most elegant of the pricing methodologies. In this approach, we begin with a set of assumptions about how the relevant variables behave. We then translate these assumptions into mathematical equations, which we use to derive a formal relationship between the input variables and the output variable (in this case the output variable is the fair option premium). The analytical model is the end result of the derivation process and takes the form of a formula or equation that ties the inputs to the output. This formula is often referred to as the solution. The beauty of analytical models is that they allow us to produce a precise valuation quickly. But there are three problems associated with analytical models.

1. The models are quite difficult to derive without an advanced knowledge of stochastic calculus.
2. In some cases, it is not possible to derive analytical solutions.
3. Analytical models, once derived, are completely inflexible. That is, they can be applied only in situations where the exact set of assumptions used to build the model hold.

Numeric models are much more flexible than analytical models. In these models, one again lays out a set of assumptions, but rather than deriving an equation that ties the inputs to the output, the model employs a finite series of steps—in an algorithmic sense—to arrive at a value. These models provide approximations of the true value rather than a precise solution. The benefit of numeric models is that they are relatively simple to build and do not require a deep understanding of stochastic calculus to appreciate or to interpret them. They are also much more flexible than analytical models. For example, one often can easily change the assumptions of the model to fit new situations. The downside is that the degree of precision is directly related to the number of computations (steps) one is prepared to employ. To get a very high degree of precision, one might have to perform tens of millions of calculations. This drawback of numeric models, however, has been rendered largely moot by the tremendous increase in the speed of microprocessors over the years and by the realization that often shortcuts can be employed with no loss of precision.

Simulation models are not as elegant as analytical models and not as fast as either analytical models or numeric models. Their great strength is that they are incredibly flexible and can be used to value derivatives that do not easily lend themselves to analytical valuation techniques or numeric valuation techniques. Indeed, the future price of any financial asset can be simulated when it is expressed in the form of an expected value.[2] In the simulation approach, we program a computer to "simulate" observations on the random variable of interest. Care must be taken to ensure that the simulated values of the variable possess the distributional properties and statistical parameters that we require.

The easiest way to see how this is done is to walk through a few exercises. We will walk through three. In the first exercise, we build a Monte Carlo simulator to value a plain vanilla call and a plain vanilla put under the same assumptions that Black and Scholes employed. Such options are written on the terminal price of the underlying. This is a good test of how accurate such a model can be. In our second exercise, we value a call and a put written on a stock's price return measured over the life of the option rather than on the stock's terminal price. The former typically trade on option exchanges and the latter typically trade in the over-the-counter options market. Finally, in our third exercise, we build a simulator to value an option written on more than one underlying asset. It is in this last case that we begin to really see the strength of simulation. As we go along, each exercise builds on the one before it.

PRICING A CLASSIC BLACK-SCHOLES OPTION

Black and Scholes made a number of assumptions. Among them, these five are important for this exercise:

1. The future price of the underlying stock is distributed lognormally.
2. The risk-free rate of interest is constant and the same for all maturities.
3. The stock's volatility is constant.
4. The option is exercisable only at the end of the option's life (i.e., it is European type).
5. The underlying stock does not pay any dividends.

Let us suppose that the current spot price of the underlying stock is \$100, the option is struck at the money (strike price = \$100), the option has three months to expiration (.25 years), the annual interest rate is 5 percent, and the underlying stock's volatility is 30 percent. For purposes of later comparison, the exact Black-Scholes value for the call is \$6.5831 and the exact Black-Scholes value for the put is \$5.3508.

In a simulation approach, we need to simulate observations on the terminal price of the stock. By *terminal price of the stock*, we mean the possible values of the stock at the point in time when the option is due to expire. These observations must have the right type of distribution (lognormal), the right mean, and the right standard deviation. Once we have simulated a possible terminal value for the stock's price at the point of the option's expiration (S_E), we can then determine the terminal value of the option. If the option is a call, its terminal value is given by $max[S_E - X, 0]$ where X denotes the strike price. If the option is a put, the terminal value of the option is given by $max[X - S_E, 0]$.

Once we have the terminal value of the option, we then discount it back to the present at the risk-free rate of interest to obtain the present value that corresponds to the terminal value. This present value is just one of an infinite number of possible present values the option might have because there are an infinite number of terminal values that the option might have. We then repeat this exercise many times. It is not unusual to run a simulator 100,000 times or more. Suppose that we run the simulator just 50,000 times. While each present value we generate is only one possible result, the average of the runs of the simulator should come very close to the true value of the option.

Let us do this simulation using the ubiquitous Microsoft product Excel®. At each step in the process, we show the relevant formula that goes into the spreadsheet cell so that you can replicate the model as we go along. There will be a box around the cell whose formula appears in the formula field.

Excel comes with one built-in random number generator, the function RAND(). The RAND() function generates random numbers having a uniform continuous distribution bounded between 0 and 1 (with a mean of 0.5). This is not the type of distribution we want, but we can get to the type of distribution we want through a series of steps starting with the RAND() function. Excel also has a function that is designed to generate an observation drawn from a standard normal distribution if you enter a cumulative probability. This is the NORMSINV() function. Since a cumulative probability must be between 0 and 1, we can generate observations

B17 f_x =NORMSINV(RAND())

	A	B	C	D	E	F	G
1							
2							
3							
4	**Stock Price=**	**100.00**					
5	**Interest Rate =**	**5.000%**					
6	**Volatility (annual)=**	**30.00%**					
7	**Call Strike=**	**100.00**					
8	**Put Strike=**	**100.00**					
9	**Time to Expiry (in years)=**	**0.2500**					
14							
15		**N(0,1)**					
16	Observation	**Z**					
17	1	**1.0573**					
18							

Exhibit 29.1 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Generating Observations

on a standard normal distribution by embedding the RAND() function in the NORMSINV() function: NORMSINV(RAND()). Call this value Z.

We illustrate the building process step by step including screen shots. Because we are employing a random number generator, at each step, the outputs that are driven by the random number generator will change automatically. This might give the impression that there is inconsistency between the different screen shots, but there is not. Please see Exhibit 29.1.

Now that we have randomly generated observations from a standard normal distribution (with a mean of 0 and a standard deviation of 1), we need to convert them to observations from a nonstandard normal distribution. Suppose that $Y \sim \mathrm{N}(\mu,\sigma)$ and define Z as $\frac{Y-\mu}{\sigma}$ so that $Z \sim \mathrm{N}(0, 1)$. This implies that $Y = Z \times \sigma + \mu$.

The nonstandard normal distribution would have a periodic mean (i.e., mean for the relevant period) equal to the expected growth rate in the stock's price over the life of the option. This is known as the drift factor or drift rate.[3] Black and Scholes discovered that, in the context of options, a stock's expected growth rate is independent of its expected return and dependent only on the risk-free rate of interest and the stock's volatility. This is counterintuitive, and is one of the key features of the Black-Scholes result, commonly known as the risk-neutrality assumption. It is important that our simulation model be consistent with the Black-Scholes framework. The drift factor (or mean) is given by:

$$\mu_{per} = \left(r - \frac{1}{2}\sigma^2\right)\tau$$

where r = risk-free rate compounded continuously
σ = annual volatility
τ = time to option expiry, measured in years (For a three-month option, $\tau = 0.25$.)

Please see Exhibit 29.2.

B11 fx =($B$5-(0.5*($B$6^2)))*$B$9

	A	B	C	D	E	F	G
1							
2							
3							
4	**Stock Price=**	**100.00**					
5	**Interest Rate =**	**5.000%**					
6	**Volatility (annual)=**	**30.00%**					
7	**Call Strike=**	**100.00**					
8	**Put Strike=**	**100.00**					
9	**Time to Expiry (in years)=**	**0.2500**					
10							
11	**periodic mean (drift factor)**	**0.125%**					

Exhibit 29.2 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Drift Factor

The standard deviation is simply the periodic volatility for the underlying stock measured over the life of the option and is simply the square root of the time to option expiry multiplied by the annual volatility:

$$\sigma_{per} = \sigma\sqrt{\tau}$$

Please see Exhibit 29.3.

Plugging these values for the mean and standard deviation into $Y = Z \times \sigma + \mu$ would give us one possible *percentage change* in the price of the underlying stock measured on the assumption of continuous compounding. Call this value Y. Please see Exhibit 29.4.

The next step is to convert this percentage change in the stock price into a terminal price for the stock. Keeping in mind that the percentage change in the price of the stock is measured on the assumption of continuous compounding, the conversion to a terminal stock price is: $S_E = exp(Y) \times S_0$.

Besides giving us the terminal price, this conversion also guarantees that the resultant terminal price will be lognormally distributed. The reason for this is straightforward. If a random variable has a lognormal distribution, then the natural

B12 fx =$B$6*($B$9^0.5)

	A	B	C	D	E	F	G
1							
2							
3							
4	**Stock Price=**	**100.00**					
5	**Interest Rate =**	**5.000%**					
6	**Volatility (annual)=**	**30.00%**					
7	**Call Strike=**	**100.00**					
8	**Put Strike=**	**100.00**					
9	**Time to Expiry (in years)=**	**0.2500**					
10							
11	**periodic mean (drift factor)**	**0.125%**					
12	**periodic standard deviation**	**15.000%**					

Exhibit 29.3 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Periodic Volatility

C17 | f_x =(B17*$B$12)+$B$11

	A	B	C	D	E	F	G
1							
2							
3							
4	**Stock Price=**	**100.00**					
5	**Interest Rate =**	**5.000%**					
6	**Volatility (annual)=**	**30.00%**					
7	**Call Strike=**	**100.00**					
8	**Put Strike=**	**100.00**					
9	**Time to Expiry (in years)=**	**0.2500**					
10							
11	**periodic mean (drift factor)**	**0.125%**					
12	**periodic standard deviation**	**15.000%**					
13							
14							
15		**N(0,1)**	**N(m,s)**				
16	Observation	**Z**	**Y**				
17	1	**2.7713**	**0.4169**				
18							

Exhibit 29.4 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Percentage Change in the Price of the Stock

log of that random variable has a normal distribution. Similarly, if a random variable has a normal distribution, the exponential of that random variable will have a lognormal distribution (since the exponential function is the inverse of the natural logarithm function). Please see Exhibit 29.5.

Once we have a simulated terminal price for the stock, we can compute the simulated terminal values for the options using the MAX() function to take the

D17 | f_x =EXP(C17)*$B$4

	A	B	C	D	E	F	G
1							
2							
3							
4	**Stock Price=**	**100.00**					
5	**Interest Rate =**	**5.000%**					
6	**Volatility (annual)=**	**30.00%**					
7	**Call Strike=**	**100.00**					
8	**Put Strike=**	**100.00**					
9	**Time to Expiry (in years)=**	**0.2500**					
10							
11	**periodic mean (drift factor)**	**0.125%**					
12	**periodic standard deviation**	**15.000%**					
13							
14							
15		**N(0,1)**	**N(m,s)**				
16	Observation	**Z**	**Y**	**S**			
17	1	**0.9552**	**0.1445**	**115.5489**			
18							

Exhibit 29.5 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Terminal Stock Price

E17 =MAX(D17-$B$7,0)

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4	Stock Price=	100.00							
5	Interest Rate =	5.000%							
6	Volatility (annual)=	30.00%							
7	Call Strike=	100.00							
8	Put Strike=	100.00							
9	Time to Expiry (in years)=	0.2500							
10									
11	periodic mean (drift factor)	0.125%							
12	periodic standard deviation	15.000%							
13									
14					Call			Put	
15		N(0,1)	N(m,s)		Terminal			Terminal	
16	Observation	Z	Y	S	Value			Value	
17	1	0.3870	0.0593	106.1090	6.1090			0.0000	
18									

Exhibit 29.6 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Option Terminal Value

maximum value. Please see Exhibit 29.6.

$$\text{Call}_{\text{E}} = max[S_E - X, 0]$$
$$\text{Put}_{\text{E}} = max[X - S_E, 0]$$

Once we have calculated the terminal values, we can simply discount them back to their present value. Under continuous compounding, the discounting equation is:

$$\text{PV} = exp(-\tau \times r) \times \text{Terminal Value}$$

Please see Exhibit 29.7.

F17 =EXP(-($B$5*$B$9))*E17

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4	Stock Price=	100.00							
5	Interest Rate =	5.000%							
6	Volatility (annual)=	30.00%							
7	Call Strike=	100.00							
8	Put Strike=	100.00							
9	Time to Expiry (in years)=	0.2500							
10									
11	periodic mean (drift factor)	0.125%							
12	periodic standard deviation	15.000%							
13									
14					Call			Put	
15		N(0,1)	N(m,s)		Terminal	Present		Terminal	Present
16	Observation	Z	Y	S	Value	Value		Value	Value
17	1	0.1388	0.0221	102.2314	2.2314	2.2037		0.0000	0.0000
18									

Exhibit 29.7 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Present Value of Option Terminal Value

Exhibit 29.8 Simulation Results

Observations	Call Price	Deviation from Black-Scholes	Put Price	Deviation from Black-Scholes
100	\$7.40	\$0.82	\$4.36	\$–0.98
1,000	\$6.92	\$0.34	\$5.13	\$–0.21
5,000	\$6.47	\$–0.11	\$5.46	\$0.12
25,000	\$6.63	\$0.05	\$5.36	\$0.02
50,000	\$6.55	\$–.03	\$5.36	\$0.02

Now that we have simulated one possible present value for the call and one possible present value for the put, we will run the simulator many times and take an average. The more times we run the simulation to obtain the average, the closer we should get to the true value of the option (essentially an application of the law of large numbers).

Simulation Results

Exhibit 29.8 presents the simulation.

In order to run the simulation 50,000, we would simply copy the formula 50,000 times and take a simple average of the simulated values. Depending on the speed of your processor, it may take a few moments to run the simulation due to the large number of calculations involved. As a side note, each worksheet in an Excel file (workbook) contains 65,536 rows, so our number of simulations is limited. But if you wanted to run more, say 100,000 simulations, you could run the 50,000 simulations twice and take an average of the two averages to replicate the results of 100,000 simulations. It would be more efficient, however, to write macros that automate the process. We will stick with 50,000 for purposes of this demonstration. Please see Exhibit 29.9.

From this simulation, the call value is \$6.7202 and the put value is \$5.3018. The exact values given earlier from the Black-Scholes' analytical model are \$6.5831 and \$5.3409 respectively. As noted, as the number of simulation runs increases, the resulting average value should converge to the values obtained from an analytical model. But there will still be some deviation due to randomness.

Price Return

Equity options that trade on exchanges are written on the stock's price. That is, the payoff at the end is dependent on the stock's price and the strike price of the option. Up until now, we have talked about equity options as though they are written on one share of stock. Of course, options typically are written on more than 1 share. Exchange traded options are usually written on 100 shares. Let us denote the number of shares by Q. Then the actual payoff on the options we have been

B13 =F50020

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4	Stock Price=	100.00							
5	Interest Rate =	5.000%							
6	Volatility (annual)=	30.00%							
7	Call Strike=	100.00							
8	Put Strike=	100.00							
9	Time to Expiry (in years)=	0.2500							
10									
11	periodic mean (drift factor)	0.125%							
12	periodic standard deviation	15.000%							
13	Simulated Call Value=	$6.7202							
14	Simulated Put Value=	$5.3018							
15									
16					Call			Put	
17		N(0,1)	N(m,s)		Terminal	Present		Terminal	Present
18	Observation	Z	Y	S	Value	Value		Value	Value
19	1	-0.8582	-0.1275	88.0317	0.0000	0.0000		11.9683	11.8196
20	2	-0.6117	-0.0905	91.3467	0.0000	0.0000		8.6533	8.5458
21	3	-0.1568	-0.0223	97.7979	0.0000	0.0000		2.2021	2.1747
22	4	1.7024	0.2566	129.2541	29.2541	28.8907		0.0000	0.0000
23	5	0.6708	0.1019	110.7236	10.7236	10.5904		0.0000	0.0000
50014	49996	1.3173	0.1988	121.9993	21.9993	21.7260		0.0000	0.0000
50015	49997	-0.4624	-0.0681	93.4164	0.0000	0.0000		6.5836	6.5018
50016	49998	0.6131	0.0932	109.7693	9.7693	9.6479		0.0000	0.0000
50017	49999	-0.3024	-0.0441	95.6855	0.0000	0.0000		4.3145	4.2609
50018	50000	1.2036	0.1818	119.9357	19.9357	19.6880		0.0000	0.0000
50019									
50020					Average	6.7202		Average	5.3018

Exhibit 29.9 Monte Carlo Simulation Analysis: Valuing Plain Vanilla Calls and Puts—Complete Spreadsheet

Note: While setting up your spreadsheet, you may wish to set Excel®'s calculation properties to manual so as to avoid automatically calculating 50,000 values each time the spreadsheet is changed (Found under Options on the Tools menu.)

discussing would be given by:

$$\text{Payoff}_{\text{call}} = max[S_E - X,\ 0] \times Q$$
$$\text{Payoff}_{\text{put}} = max[X - S_E,\ 0] \times Q$$

Unlike exchange-traded options, equity options that trade over the counter are often written on price return rather than price. That is, the payoff at the end of the life of the option is given by $max[PR - XR, 0] \times NP$ for calls and $\max[\text{XR} - \text{PR}, 0] \times$ NP for puts. Here, *PR* is the price return, defined as the percentage change in the price of the stock (or stock index) over the life of the option, *XR* is the strike rate, and *NP* is the notional principal. In the case of an at-the-money call, this would be $max[PR - 0\%, 0] \times NP$.

To see that option payoffs defined in terms of terminal stock price and those defined in terms of price return are really the same thing, we demonstrate that we can easily move from one to the other:

$$\text{Payoff}_{\text{call}} = max[S_E - X,\ 0] \times Q$$

If we multiply by $\frac{S_0}{S_0}$:

$$
\begin{aligned}
&= max\,[S_E - X,\ 0] \times Q \times \frac{S_0}{S_0} \\
S_E &= \frac{max\,[S_E - X,\ 0]}{S_0} \times Q \times S_0 \\
&= max\left[1 + PR - \left(1 + \frac{X}{S_0}\right),\ \frac{0}{S_0}\right] \times NP \qquad \text{where } PR = \frac{S_E}{S_0} - 1 \\
&= max\,[PR - XR,\ 0] \times NP \qquad \text{where } XR = \frac{X}{S_0} - 1
\end{aligned}
$$

We can easily adapt the simulation model we built previously to price this option. In this exercise, the price of the option is quoted as a percentage of the notional principal on which the option is written. That is, we work in terms of \$1 of notional principal. Instead of inputting strike prices, we would input the strike rates for the call and the put. The price return can then be calculated by taking the terminal stock price and dividing it by the beginning stock price. The terminal values for the call and put are then calculated by $max[PR - XR, 0]$ and $max[XR - PR, 0]$, respectively. The final step is to discount to obtain the present value. This is done the same way as for the earlier model. In this particular case, with 50,000 simulations in the run, we obtained a price for the call of 6.542 percent and a price for the put of 5.786 percent. In both cases, these are interpreted as percentages of the notional principal on which the options are written. For example, if the client wanted to buy the call on \$1,000,000 of notional principal, the price of the option would be \$65,420. Please see Exhibit 29.10.

E19 =D19/$B$4-1

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4	Stock Price=	100.00									
5	Interest Rate =	5.000%									
6	Volatility (annual)=	30.00%									
7	Call Strike Rate=	0.00%									
8	Put Strike Rate=	0.00%									
9	Time to Expiry (in years)=	0.2500									
10											
11	periodic mean (drift factor)	0.125%									
12	periodic standard deviation	15.000%									
13	Simulated Call Value=	6.542%									
14	Simulated Put Value=	5.786%									
15											
16							Call			Put	
17		N(0,1)	N(m,s)		Price		Terminal	Present		Terminal	Present
18	Observation	Z	Y	S	Return		Value	Value		Value	Value
19	1	-1.7797	-0.2657	76.6664	-23.33%		0.00%	0.00%		23.33%	23.04%
20	2	0.9604	0.1453	115.6390	15.64%		15.64%	15.44%		0.00%	0.00%
21	3	2.0371	0.3068	135.9089	35.91%		35.91%	35.46%		0.00%	0.00%
22	4	0.6349	0.0965	110.1291	10.13%		10.13%	10.00%		0.00%	0.00%
23	5	0.6456	0.0981	110.3069	10.31%		10.31%	10.18%		0.00%	0.00%
24	6	-1.6119	-0.2405	78.6210	-21.38%		0.00%	0.00%		21.38%	21.11%
25	7	0.2071	0.0323	103.2843	3.28%		3.28%	3.24%		0.00%	0.00%
26	8	-0.3437	-0.0503	95.0942	-4.91%		0.00%	0.00%		4.91%	4.84%
27	9	-0.8433	-0.1252	88.2284	-11.77%		0.00%	0.00%		11.77%	11.63%
28	10	1.0942	0.1654	117.9841	17.98%		17.98%	17.76%		0.00%	0.00%

Exhibit 29.10 Monte Carlo Simulation Analysis: Valuing Over-the-Counter Calls and Puts (Price Return)

You will have noticed that, with the proper interpretation, options written on an asset's price and options written on an asset's price return are really the same thing. However, a key advantage of modeling an option on price returns is that it more easily lends itself to pricing options written on multiple underlyings. We take these up next in the context of rainbow options. The term *rainbow option* is an industry term for options with more than one underlying asset. If there are two underlying assets, often they are called two-color rainbows. If there are three underlying assets, they are called three-color rainbows, and so forth. These should not be confused with index options, which have only one underlying: the index. Rainbows typically are written on price returns and fall within the realm of exotic options. It is in the valuation of exotic options that simulation shows its real strength.

PRICING A RAINBOW OPTION

Rainbow options come in a number of different varieties, and each has its own purposes. We illustrate them with two underlyings (but there can be any number of underlyings). Consider just a few of the possible payoff structures. Rather than give them formal names, we simply refer to them as Types 1 through 6.

Type 1: Payoff $= max[PR_1, PR_2]$
Type 2: Payoff $= max[PR_1, PR_2, 0]$
Type 3: Payoff $= max[max(PR_1, PR_2) - XR, 0]$
Type 4: Payoff $= max[max(PR_1 - XR_1, PR_2 - XR_2), 0]$
Type 5: Payoff $= max[min(PR_1, PR_2) - XR, 0]$
Type 6: Payoff $= max[min(PR_1 - XR_1, PR_2 - XR_2), 0]$

Type 1, often called a best-of or better-of option, pays off based on the best performing of two underlyings. But if both underlyings produce a negative price return, the payoff is actually negative (not zero as in a conventional option). According to Smithson (1998), a popular combination in the early 1990s was a Type 1 two-color rainbow option based on the performance of a stock market index and a bond market index. An investor struggling with the decision of whether to buy stocks or bonds could buy a Type 1 option on which the payoff is determined by which of the two asset classes performs better.

Type 2 is similar to Type 1, except that it guarantees that the payoff will never be negative. Notice that Type 1 and Type 2 do not provide for a strike rate. Type 3 does allow for a strike rate, but the same strike rate applies to both underlyings. If the strike rate is set to zero, then Type 3 collapses to Type 2. Type 4 allows for more than one strike rate; that is, a different strike rate can be applied to each of the underlyings (often called a dual-strike rainbow). However, if the two strikes happen to be the same, Type 4 collapses to Type 3. Types 5 and 6 are analogous to Types 3 and 4 except that Types 5 and 6 pay off based on the worst performer, rather than the best performer. All six of these options can be thought of as calls. We will build a simulation model to price up Types 4 and 6.

Suppose that we are going to write a single option on two U.S. stocks. Let us call these stocks ABC and XYZ. In one case, the payoff would be based on the better performer (Type 4), and in the other case, the payoff would be based on the worst performer (Type 6). For simplicity, assume that the option is written at-the-money with respect to both of the underlying stocks and neither stock pays dividends. The payoff, per \$1 of notional principal, for a call of Type 4 at its expiry would be $max[max(PR_{ABC} - 0\%, PR_{XYZ} - 0\%), 0]$. And the payoff for a call of Type 6 at its expiry would be $max[min(PR_{ABC} - 0\%, PR_{XYZ} - 0\%), 0]$.

What complicates things now, relative to our previous example of an option written on a single price return, is that we have an additional value driver. This additional value driver is the *correlation* of the price returns on the two stocks. How the correlation will impact the value of a rainbow option depends on the nature of the payoff function. This correlation has to be included in the simulated terminal stock values.

To have a concrete example to work with, let us suppose that ABC is currently priced at \$100 a share and has an annual volatility of 25 percent, and XYZ is currently priced at \$50 a share and has an annual volatility of 45 percent. The options both have three months to expiry ($\tau = 0.25$ years). As already noted, for simplicity, we assume that the options are written at-the-money with respect to both stocks so that the strike rates XR_1 and XR_2 are both zero. The annual interest rate is 5 percent, and the degree of correlation between the two stocks' returns is 0.65.

We begin as we did in the last model, except that instead of generating observations on one standard normal random variable, we have to simultaneously generate observations on two separate standard normal random variables. In both cases, we use the NORMSINV(RAND()) function. Denote these Z_1 and Z_2. These two variables are uncorrelated with each other. They must therefore be "adjusted" to bring in the correlation. This is a simple procedure that employs a well-known statistical relationship. Define two new random variables R_1 and R_2 and let $R_1 = Z_1$ and $R_2 = \rho Z_1 + Z_2\sqrt{1-\rho^2}$, where ρ denotes the degree of correlation between the two stocks' returns. R_1 and R_2 are now standard normal random variables (with means of 0 and standard deviations of 1 *and* they have the desired degree of correlation. Please see Exhibit 29.11.

We now have to convert these standard normal distributions to appropriate nonstandard normal distributions. We denote these as Y_1 and Y_2. Y_1 and Y_2 will be calculated from R_1 and R_2 by incorporating an appropriate mean and standard deviation for each. The mean, standard deviation, terminal stock price, and price return for each stock are calculated in the same manner as we did in the last exercise. As before, we will run the simulator 50,000 times. Please see Exhibit 29.12.

The terminal value equation for the Type 4 call is: $max[max(PR_1 - XR_1, PR_2 - XR_2), 0]$. The terminal value equation for the Type 6 call is $max[min(PR_1 - XR_1, PR_2 - XR_2), 0]$. Please see Exhibits 29.13 and 29.14.

Once the terminal values have been calculated, we can use the continuous compounding discounting equation to determine each possible present value for the Type 4 option and each possible present value for the Type 6 option. We then run many simulations and take an average. Notice that when the returns have a correlation of 0.65, Type 4 options are worth about 11.27 percent of the notional principal and Type 6 options are worth about 3.70 percent of the notional principal. (These values will vary a bit each time the simulator is used.) From your own simulation, you should see that changing the correlation

E22 f_x =($B$10*B22)+(C22*SQRT((1-($B$10^2))))

	A	B	C	D	E	Q	R	S
1								
2								
3		Stock 1	Stock 2					
4	Stock Price=	100.00	50.00					
5	Interest Rate =	5.000%						
6	Volatility (annual)=	25.00%	45.00%					
7	Type 4 Strike Rate=	0.00%	0.00%					
8	Type 6 Strike Rate=	0.00%	0.00%					
9	Time to Expiry (in years)=	0.2500						
10	Correlation=	0.6500						
11								
12	periodic mean (drift factor)	0.469%	-1.281%					
13	periodic standard deviation	12.500%	22.500%					
14								
19								
20		N(0,1)	N(0,1)	N(0,1)	N(0,1)			
21	Observation	Z1	Z2	R1	R2			
22	1	-1.5312	0.0223	-1.5312	-0.9784			
23	2	-1.3815	1.6232	-1.3815	0.3356			
24	3	1.1933	-0.8405	1.1933	0.1369			
25	4	0.2509	-0.3084	0.2509	-0.0713			
26	5	-0.4047	-0.3212	-0.4047	-0.5072			
50017	49996	-0.5600	-0.6834	-0.5600	-0.8833			
50018	49997	0.4625	-0.4862	0.4625	-0.0689			
50019	49998	-1.1020	-0.3230	-1.1020	-0.9617			
50020	49999	2.5944	0.4642	2.5944	2.0391			
50021	50000	-0.7516	-0.5744	-0.7516	-0.9250			

Exhibit 29.11 Monte Carlo Simulation Analysis: Valuing a Rainbow Option of Type 4 and Type 6—Generating Observations

K22 f_x =I22/$C$4-1

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		Stock 1	Stock 2								
4	Stock Price=	100.00	50.00								
5	Interest Rate =	5.000%									
6	Volatility (annual)=	25.00%	45.00%								
7	Type 4 Strike Rate=	0.00%	0.00%								
8	Type 6 Strike Rate=	0.00%	0.00%								
9	Time to Expiry (in years)=	0.2500									
10	Correlation=	0.6500									
11											
12	periodic mean (drift factor)	0.469%	-1.281%								
13	periodic standard deviation	12.500%	22.500%								
14											
19											
20		N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(m,s)	N(m,s)			Price	Price
21	Observation	Z1	Z2	R1	R2	Y1	Y2	S1	S2	Return1	Return2
22	1	-1.5312	0.0223	-1.5312	-0.9784	-0.1867	-0.2329	82.9678	39.6099	-17.03%	-20.78%
23	2	-1.3815	1.6232	-1.3815	0.3356	-0.1680	0.0627	84.5359	53.2354	-15.46%	6.47%
24	3	1.1933	-0.8405	1.1933	0.1369	0.1538	0.0180	116.6310	50.9079	16.63%	1.82%
25	4	0.2509	-0.3084	0.2509	-0.0713	0.0360	-0.0289	103.6703	48.5780	3.67%	-2.84%
26	5	-0.4047	-0.3212	-0.4047	-0.5072	-0.0459	-0.1269	95.5139	44.0400	-4.49%	-11.92%
50017	49996	-0.5600	-0.6834	-0.5600	-0.8833	-0.0653	-0.2116	93.6780	40.4660	-6.32%	-19.07%
50018	49997	0.4625	-0.4862	0.4625	-0.0689	0.0625	-0.0283	106.4498	48.6046	6.45%	-2.79%
50019	49998	-1.1020	-0.3230	-1.1020	-0.9617	-0.1331	-0.2292	87.5413	39.7586	-12.46%	-20.48%
50020	49999	2.5944	0.4642	2.5944	2.0391	0.3290	0.4460	138.9558	78.1017	38.96%	56.20%
50021	50000	-0.7516	-0.5744	-0.7516	-0.9250	-0.0893	-0.2209	91.4609	40.0882	-8.54%	-19.82%

Exhibit 29.12 Monte Carlo Simulation Analysis: Valuing a Rainbow Option of Type 4 and Type 6—Converting to Non-Standard Normal Distributions

L22 =MAX(MAX(J22-$B$7,K22-$C$7),0)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3		Stock 1	Stock 2										
4	Stock Price=	100.00	50.00										
5	Interest Rate =	5.000%											
6	Volatility (annual)=	25.00%	45.00%										
7	Type 4 Strike Rate=	0.00%	0.00%										
8	Type 6 Strike Rate=	0.00%	0.00%										
9	Time to Expiry (in years)=	0.2500											
10	Correlation=	0.6500											
11													
12	periodic mean (drift factor)	0.469%	-1.281%										
13	periodic standard deviation	12.500%	22.500%										
14													
15	Simulated Type 4 Value=	11.266%											
16	Simulated Type 6 Value=	3.701%											
17													
18													
19												Type 4	
20		N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(m,s)	N(m,s)			Price	Price	Terminal	Present
21	Observation	Z1	Z2	R1	R2	Y1	Y2	S1	S2	Return1	Return2	Value	Value
22	1	-1.5312	0.0223	-1.5312	-0.9784	-0.1867	-0.2329	82.9678	39.6099	-17.03%	-20.78%	0.00%	0.00%
23	2	-1.3815	1.6232	-1.3815	0.3356	-0.1680	0.0627	84.5359	53.2354	-15.46%	6.47%	6.47%	6.39%
24	3	1.1933	-0.8405	1.1933	0.1369	0.1538	0.0180	116.6310	50.9079	16.63%	1.82%	16.63%	16.42%
25	4	0.2509	-0.3084	0.2509	-0.0713	0.0360	-0.0289	103.6703	48.5780	3.67%	-2.84%	3.67%	3.62%
26	5	-0.4047	-0.3212	-0.4047	-0.5072	-0.0459	-0.1269	95.5139	44.0400	-4.49%	-11.92%	0.00%	0.00%
50017	49996	-0.5600	-0.6834	-0.5600	-0.8833	-0.0653	-0.2116	93.6780	40.4660	-6.32%	-19.07%	0.00%	0.00%
50018	49997	0.4625	-0.4862	0.4625	-0.0689	0.0625	-0.0283	106.4498	48.6046	6.45%	-2.79%	6.45%	6.37%
50019	49998	-1.1020	-0.3230	-1.1020	-0.9617	-0.1331	-0.2292	87.5413	39.7586	-12.46%	-20.48%	0.00%	0.00%
50020	49999	2.5944	0.4642	2.5944	2.0391	0.3290	0.4460	138.9558	78.1017	38.96%	56.20%	56.20%	55.51%
50021	50000	-0.7516	-0.5744	-0.7516	-0.9250	-0.0893	-0.2209	91.4609	40.0882	-8.54%	-19.82%	0.00%	0.00%
50022												Average	11.27%

Exhibit 29.13 Monte Carlo Simulation Analysis: Valuing a Rainbow Option of Type 4 and Type 6—Option Terminal Value Type 4

O22 =MAX(MIN(J22-$B$8,K22-$C$8),0)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3		Stock 1	Stock 2													
4	Stock Price=	100.00	50.00													
5	Interest Rate =	5.000%														
6	Volatility (annual)=	25.00%	45.00%													
7	Type 4 Strike Rate=	0.00%	0.00%													
8	Type 6 Strike Rate=	0.00%	0.00%													
9	Time to Expiry (in years)=	0.2500														
10	Correlation=	0.6500														
11																
12	periodic mean (drift factor)	0.469%	-1.281%													
13	periodic standard deviation	12.500%	22.500%													
14																
15	Simulated Type 4 Value=	11.266%														
16	Simulated Type 6 Value=	3.701%														
17																
18																
19												Type 4			Type 6	
20		N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(m,s)	N(m,s)			Price	Price	Terminal	Present		Terminal	Present
21	Observation	Z1	Z2	R1	R2	Y1	Y2	S1	S2	Return1	Return2	Value	Value		Value	Value
22	1	-1.5312	0.0223	-1.5312	-0.9784	-0.1867	-0.2329	82.9678	39.6099	-17.03%	-20.78%	0.00%	0.00%		0.00%	0.00%
23	2	-1.3815	1.6232	-1.3815	0.3356	-0.1680	0.0627	84.5359	53.2354	-15.46%	6.47%	6.47%	6.39%		0.00%	0.00%
24	3	1.1933	-0.8405	1.1933	0.1369	0.1538	0.0180	116.6310	50.9079	16.63%	1.82%	16.63%	16.42%		1.82%	1.79%
25	4	0.2509	-0.3084	0.2509	-0.0713	0.0360	-0.0289	103.6703	48.5780	3.67%	-2.84%	3.67%	3.62%		0.00%	0.00%
26	5	-0.4047	-0.3212	-0.4047	-0.5072	-0.0459	-0.1269	95.5139	44.0400	-4.49%	-11.92%	0.00%	0.00%		0.00%	0.00%
50017	49996	-0.5600	-0.6834	-0.5600	-0.8833	-0.0653	-0.2116	93.6780	40.4660	-6.32%	-19.07%	0.00%	0.00%		0.00%	0.00%
50018	49997	0.4625	-0.4862	0.4625	-0.0689	0.0625	-0.0283	106.4498	48.6046	6.45%	-2.79%	6.45%	6.37%		0.00%	0.00%
50019	49998	-1.1020	-0.3230	-1.1020	-0.9617	-0.1331	-0.2292	87.5413	39.7586	-12.46%	-20.48%	0.00%	0.00%		0.00%	0.00%
50020	49999	2.5944	0.4642	2.5944	2.0391	0.3290	0.4460	138.9558	78.1017	38.96%	56.20%	56.20%	55.51%		38.96%	38.47%
50021	50000	-0.7516	-0.5744	-0.7516	-0.9250	-0.0893	-0.2209	91.4609	40.0882	-8.54%	-19.82%	0.00%	0.00%		0.00%	0.00%
50022												Average	11.27%		Average	3.70%

Exhibit 29.14 Monte Carlo Simulation Analysis: Valuing a Rainbow Option of Type 4 and Type 6—Option Terminal Value Type 6

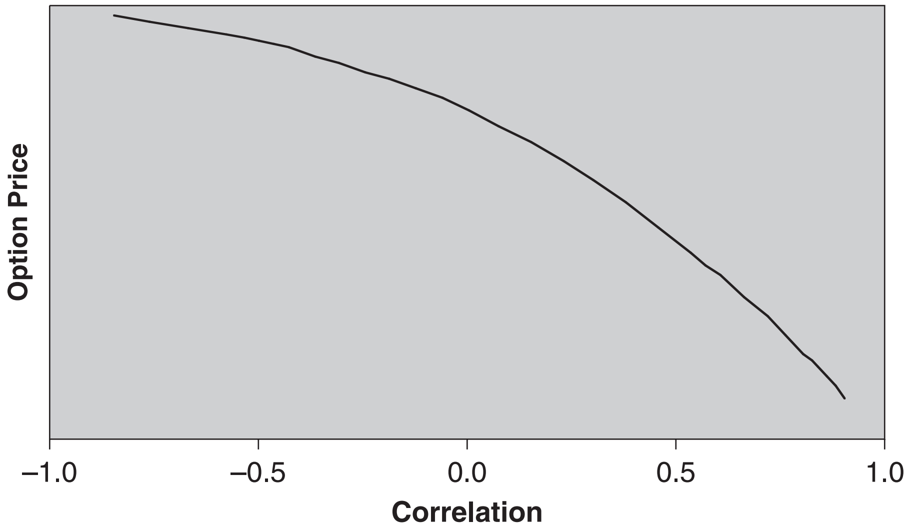


Exhibit 29.15 Value of a Type 4 Rainbow Option as a Function of Correlation

between the stocks will change the results. Clearly, correlation must be factored into the valuation methodology, and Monte Carlo simulation makes this relatively easy to do.

As demonstrated by Kolb (2007), all other things being equal, the higher the correlation between the returns of two assets, the lower the value of a Type 4 rainbow. For a Type 6 rainbow, the opposite is true. Please see Exhibit 29.15.

This model can easily be adapted to handle any of the various types of rainbows we described at the beginning of this section, other types of rainbows, and rainbows of as many "colors" as we like.

Given the incredible flexibility of Monte Carlo simulation, it is no wonder that it has become a tool of choice for valuing complex derivatives.

ENDNOTES

1. Many people consider simulation models to be a subset of numeric models. I believe that they are sufficiently different to be in a class by themselves, which is typically the way they are viewed by the financial services firms that use simulation models so effectively.
2. Elaboration of this point can be found in Briys et al. (1998).
3. The drift factor is the average increase per unit of time in a stochastic variable. Derivation of the drift factor formula is beyond the scope of this demonstration. For those who would like to pursue it more deeply, see Hull (2005), Chapter 12.

REFERENCES

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, no. 3: 637–654.

Briys, E., M. Bellalah, H. M. Mai, and F. de Varenne. 1998. *Options, Futures and Exotic Derivatives, Theory, Application and Practice.* New York: John Wiley & Sons.

Cox, J. C., S. A. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach," *Journal of Financial Economics* 7: 229–263.

Hull, J. 2005. *Options, Futures & Other Derivatives*, 6th ed. Upper Saddle River, NJ: Prentice Hall.

Kolb, R. W. 2007. *Futures, Options, & Swaps*, 5th edition. Malden, MA: Blackwell Publishing.

McLeish, D. L. 2005. *Monte Carlo Simulation and Finance.* Hoboken, NJ: John Wiley & Sons.

Smithson, C. W. 1998. *Managing Financial Risk: A Guide to Derivative Products, Financial Engineering, and Value Maximization*, 3rd ed. New York: McGraw-Hill.

ABOUT THE AUTHOR

Cara Marshall is lecturer of economics/finance at Queens College, City University of New York. Her research interests focus on derivatives and financial engineering as well as behavioral and experimental methods in finance. Her PhD dissertation examined the pricing of volatility on U.S. options exchanges. The first half of the study concerned the efficiency with which the market prices index volatility (as derivable from index options) relative to the individual volatilities of the index's components (as derivable from equity options). The second half of the study compared the implied volatilities of individual stocks derived in two different ways. Prior to academia, Dr. Marshall worked in Internet engineering, developing Web sites and an online platform for online course delivery. Over the years, she has served as a training consultant to several investment banks. In this role, she taught financial modeling to bank employees in New York, London, and Singapore.

CHAPTER 30

Valuing Derivatives Using Finite Difference Methods

CRAIG PIRRONG
Professor of Finance, Director, Global Energy Management Institute,
Bauer College of Business, University of Houston

INTRODUCTION

The valuation of options and other contingent claims typically requires the use of computational methods. A variety of computational methods are available. Of these, finite difference methods are among the most well known, most reliable, and most flexible. Finite difference techniques approximate the solution to the partial differential equation that Black, Scholes, and Merton demonstrated that any contingent claim must satisfy when the assumptions they made about stock price dynamics hold.[1] The user of the method can control the accuracy of the approximation, with more accurate approximations requiring greater use of computational resources. Finite difference methods are an essential tool in any derivatives modeler's kit because they can be adapted to a wide variety of problems, including the valuation of options that permit early exercise and many exotic derivatives.

This chapter discusses the use of finite difference methods to value contingent claims. I first discuss several basic valuation methods—explicit integration, Monte Carlo integration, and finite difference methods—to put the latter in a more general context. I then set out the common finite difference methods—explicit, implicit, and Crank-Nicolson—used to solve for one-dimensional problems. I then discuss the use of finite difference methods in multidimensional problems, such as options on multiple underlying assets. Finally, I examine the pros and cons of finite difference methods relative to alternative numerical techniques.

AN OVERVIEW

Two basic theoretical frameworks—Martingale methods and the partial differential equation (PDE) approach—motivate two different but equivalent ways of valuing derivative securities. An understanding of these alternative approaches sheds light on the relations between alternative numerical techniques for pricing contingent claims.

Martingale methods demonstrate that the value of any contingent claim is the expected present value of the claim's cash flows, where the expectation is taken under the "equivalent" probability measure; one example of this equivalent measure is the "risk-neutral" measure where risky securities earn the risk-free rate of return and all cash flows can be discounted at the risk-free rate. When the market is complete (in the sense that the payoffs of any contingent claim can be replicated through a dynamic trading strategy), this equivalent measure is unique.

The Martingale approach motivates two ways of valuing a contigent claim. The first is to calculate the expectation by explicitly integrating the payoff to an option. For instance, when the stock price follows a geometric Brownian motion, one can solve:

$$C(S_t, K, \sigma, T, t) = \int_{-\infty}^{W^*} \left[S_t^{(r-.5\sigma^2)(T-t)-W\sigma\sqrt{T-t}W} - K \right] \frac{e^{-.5W^2}}{\sqrt{2\Pi}} dW$$

to determine the value of a call. The term $S_t^{(r-.5\sigma^2)(T-t)-W\sigma\sqrt{T-t}W}$ is the value of the stock price at expiration date T if the unexpected portion of stock's return between t and expiration equals $W\sigma\sqrt{T-t}$, where W is a standard normal deviate, and K is the strike price of the option. Hence, the term in brackets is the payoff of the option conditional on a realization of the return shock W, and the normal density term that multiplies the payoff is the probability of observing such a realization. In this expression

$$W^* = \frac{\ln \frac{S_t}{K} + (r - .5\sigma^2)(T - t)}{\sigma\sqrt{T - t}}$$

is the critical value of the random normal shock to the stock's return that the option is just at the money at expiration. Simplifying the integral in the expression for $C(S_t, K, \sigma, T, t)$ produces the familiar Black-Scholes equation for a call.

The integration method is tractible only when the option has a payoff at a single date (such as at expiration) or at most at a small number of dates; cash flows at N dates require the evaluation of order N integrals, and the computational cost grows geometrically with the order of integration. The integration approach is not generally practical for American options, especially American puts, where the option could potentially have a payoff at any date.

The second valuation method motivated by the Martingale approach is Monte Carlo integration. In this approach, one simulates a large number of possible values of the stock price at expiration (or a large number of paths of the stock price for path-dependent derivatives) and approximates the relevant integral as the average of the present value of option payoffs across all the simulations.

Monte Carlo is a flexible, computationally intensive approach that has some advantages but some disadvantages as well. For instance, since Monte Carlo is a simulation method, valuation estimates are subject to sampling error. These sampling errors can be reduced by increasing the number of simulation runs, and hence computational cost, or through the use of various error control techniques. Moreover, calcuation of hedging parameters—"the Greeks"—is often difficult in Monte Carlo. Furthermore, Monte Carlo often runs into difficulties when estimating the value of path-dependent instruments, such as barrier options, especially if the every point along the stock price's path potentially affects value (as is the case for barrier options with continuously monitored barriers). In addition, although recent advances allow the use of Monte Carlo to value American options, these methods require substantially higher computational costs to achieve a level of accuracy comparable with that incurred when valuing European options using Monte Carlo.

Finite difference methods are a different numerical technique that offers some benefits over explicit integration and Monte Carlo. The finite difference method traces its roots to an earlier approach to valuing derivatives, the PDE technique devised by Black, Scholes, and Merton (BSM). Specifically, BSM showed that if a stock's price follows a geometric Brownian motion (GBM), the price V of a contingent claim on the stock must satisfy the following partial differential equation (PDE):

$$rV = V_t + (r - \delta)S_t V_S + .5\sigma^2 S_t^2 V_{SS} \tag{30.1}$$

A standard log transformation converts Equation 30.1 into a PDE with constant coefficients that is often more convenient to solve:

$$rV = V_t + (r - \delta - .5\sigma^2)V_Z + .5\sigma^2 V_{ZZ} \tag{30.2}$$

where $Z =$ natural logarithm of the stock price
$r =$ riskless interest rate (assumed constant)
$\delta =$ dividend yield on the stock
$S_t =$ stock price at time t
$\sigma =$ volatility of the stock price
subscripts on $V =$ partial derivatives

This is a one-size-fits-all equation that characterizes all contingent claims on the stock. Solution of this second-order parabolic equation subject to the boundary conditions unique to a particular derivative provides the no-arbitrage price of that contingent claim as a function of the stock price, time to expiration, the interest rate, and the volatility of the stock.

In some cases, notably for European vanilla options, there are closed-form solutions to Equations 30.1 or 30.2; these are identical to those produced by explicit integration. In other cases, however, there are no closed-form solutions. For instance, there is no closed-form solution for the value of an American put on the stock. Numerical techniques must be employed to determine the value of these claims. Finite difference methods represent the most flexible, and most commonly employed, method to solve Equation 30.1 in cases where a closed-form solution does not exist.[2]

The Feynman-Kac theorem provides mathematical link between finite difference methods and the Martingale-inspired, integration-based techniques. This theorem states that if a function V gives the expected present value (under the equivalent measure) of a contingent claim's cash flows, then V must solve the PDE Equation 30.1 or 30.2. Rather than estimating an integral, finite difference methods determine the derivative's value by solving the PDE. However, as Feynman-Kac shows, the integration and PDE approaches are isomorphic.

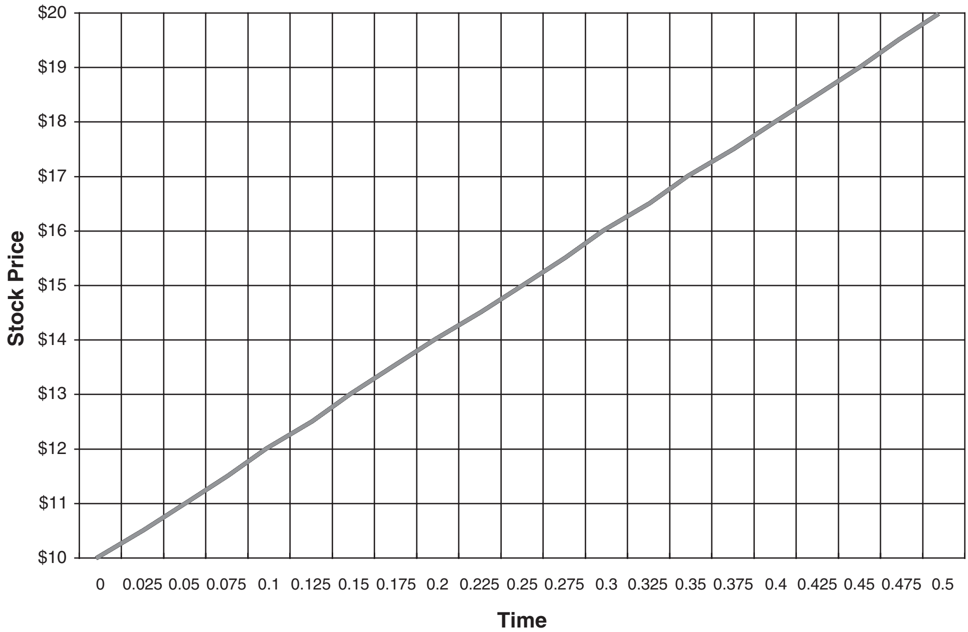


Exhibit 30.1 Valuation Grid

Determination of V from Equation 30.1 or 30.2 involves the solution for an entire *function* of the stock price and time.[3] Solving for this function exactly typically is impossible, so instead the function is approximated at a finite number of stock price and time points, which constitute a mesh or grid or lattice of such points; as a result, finite difference methods are also referred to as mesh, grid, or lattice methods.

In essence, the valuation proceeds by dividing time between the present and expiration, and the stock price, into discrete increments. Exhibit 30.1 illustrates such a valuation grid. Time is depicted on the horizontal axis, with expiration at the far right of the grid and the valuation date at the left. The stock price is depicted on the vertical axis. Finite difference methods determine the option value at each point in this grid.

Working from expiration back to the present, the method solves Equation 30.1 or 30.2 at each of the discrete time and stock price steps. As the phrase *finite difference* suggests, this approach approximates the partial derivatives in Equation 30.1 or 30.2 with differences on this mesh or grid. As a result of these approximations, Equation 30.1 or 30.2 is transformed into a set of linear equations that can be solved to determine a claim's value.

There are a variety of finite difference methods, which differ depending on how these derivatives are estimated. Explicit methods estimate the relevant stock price derivatives in V at time t (corresponding to one of time steps in the grid) in Equation 30.1 or 30.2 using the value of V at *future* dates. Since valuation proceeds from expiration back to the present, at any time step the values of V at future time steps (i.e., times closer to expiration) are already known, so the derivatives are also known. Thus, explicit methods are easy to implement because they require no solution of simultaneous equations. However, as will be seen, this method is potentially unstable, and controlling instability inflates computational costs.

The well-known binomial model is one example of an explicit finite difference method. The trinomial model is another.

Implicit methods estimate V_S and V_{SS} in Equation 30.1 (or V_Z and V_{ZZ} in Equation 30.2) for time t using the values of V at t. This requires the solution of a system of simultaneous equations, which increases computational cost. However, the implicit method is unconditionally stable.

Crank-Nicolson methods essentially combine explicit and implicit approximations of V_S and V_{SS}. This method is unconditionally stable but exhibits some undesirable spurious fluctuations in value, especially around the money.

All finite difference methods must utilize information about the value of the contingent claims at extreme values of the stock price (i.e., $S_t = \infty$ and $S_t = 0$). These "boundary conditions" are derived from the fundamental features of the claim being valued and are necessary in order to ensure solution of the linear equations implied by Equation 30.1 or 30.2 and the discretization.

All finite difference methods can be used to value options with early exercise features; this is one of their main virtues. Moreover, these methods also permit accurate calculations of the Greeks. In addition, with some additional programming work and computational cost, these methods also can value some path dependent options. These methods are particularly useful in valuing barrier options because the barriers can be incorporated into the boundary conditions needed to solve the equations implied by the relevant PDE.

In brief, finite difference methods are just one of many numerical methods to value a contingent claim. All methods should give the same answer, and the decision of which method to use depends on the nature of the instrument being valued. Finite difference methods are best for options with early exercise and some exotic claims (such as barrier options), and when accurate information on the Greeks is necessary. I now turn attention to the details of the implementation of this approach.

BASIC METHODS

Finite difference methods commence by discretizing the stock price and time dimensions to create a two-dimensional valuation grid. It is conventional to divide time between the valuation date and the expiration date of the contingent claim into equal increments δt in length. In what follows, I assume there are N such intervals between the valuation and expiration dates, and hence $N + 1$ time points in the grid. Time step $j = 1$ corresponds to the valuation date, and time step $j = N + 1$ corresponds to expiration. Similarly, it is conventional to divide the log stock price into equal-size increments δZ in length. Since Z is unbounded above and below (the stock price ranges between zero and infinity, so Z lies between plus and minus infinity), the log stock price interval must be truncated to $[\underline{Z}, \overline{Z}]$. This truncation introduces some error into the estimate of the claim value. There are I log stock price intervals and $I + 1$ log stock price points in the grid.

The modeler chooses δt and δZ and faces a trade-off when doing so. Smaller time and log stock price intervals improve the accuracy of the numerical estimate of the claim value but increase computational costs. Similarly, errors are smaller with a larger range $[\underline{Z}, \overline{Z}]$, but increasing this range increases computational costs.

The essence of the finite difference method is to approximate the partial derivatives in Equation 30.2. For all standard methods, at time step j and log stock price step i the time partial derivative V_t is approximated:

$$V_t \approx \frac{V_i^{j+i} - V_i^j}{\delta t}$$

where superscript = time step where V is estimated,
subscript = log stock price step of the approximation

The three basic finite difference methods differ in their approximation of the log stock price partials. At time step j, the *explicit* method uses the values of V at the next time step, $j+1$, to approximate the log stock price partials at stock price step i and time step j as shown:

$$V_Z \approx \frac{V_{i+i}^{j+1} - V_{i-1}^{j+1}}{2\delta Z}$$

$$V_{ZZ} \approx \frac{V_{i+i}^{j+1} - 2V_i^{j+1} + V_{i-1}^{j+1}}{\delta Z^2}$$

The *implicit* method uses the values of V at the contemporaneous time step j to approximate the partials at log stock price step i and time step j as shown:

$$V_Z \approx \frac{V_{i+i}^{j} - V_{i-1}^{j}}{2\delta Z}$$

$$V_{ZZ} \approx \frac{V_{i+i}^{j} - 2V_i^{j} + V_{i-1}^{j}}{\delta Z^2}$$

The Crank-Nicolson method averages the approximations from the implicit and explicit methods:

$$V_Z \approx .5\frac{V_{i+i}^{j+1} - V_{i-1}^{j+1}}{2\delta Z} + .5\frac{V_{i+i}^{j} - V_{i-1}^{j}}{2\delta Z}$$

$$V_{ZZ} \approx .5\frac{V_{i+i}^{j+1} - 2V_i^{j+1} + V_{i-1}^{j+1}}{\delta Z^2} + .5\frac{V_{i+i}^{j} - 2V_i^{j} + V_{i-1}^{j}}{\delta Z^2}$$

Substituting these approximations of the partial derivatives into Equation 30.2 produces systems of linear equations. For the explicit method, at each interior time $(j = 1, \ldots, N)$ and stock price step $(i = 2, \ldots, I)$ of the grid, after substitution and some rearrangement, the relevant equation is:

$$V_i^j = A_E V_{i+1}^{j+1} + B_E V_i^{j+1} + C_E V_{i-1}^{j+1} \tag{30.3}$$

where

$$A_E = \frac{.5\sigma^2\delta t}{\delta Z^2} + \frac{.5(r - \delta - .5\sigma^2)\delta t}{\delta Z}$$
$$B_E = 1 + r\delta t - \frac{\sigma^2\delta t}{\delta Z^2}$$

and

$$C_E = \frac{.5\sigma^2\delta t}{\delta Z^2} - \frac{.5(r - \delta - .5\sigma^2)\delta t}{\delta Z}$$

Note that the value of the option at time step j depends on three values of the option at the next time step. Hence, this method is sometimes referred to as the trinomial method.[4]

In the explicit method, one first fills in the values of the option at expiration (i.e., at time step $N + 1$). For a put with strike price K, for instance, the payoff at expiration at log price step i is:

$$V_i^{I+1} = \max\left[0,\ K - e^{Z_i}\right]$$

Then one proceeds to time step N (one step prior to expiration), and solves 30.3 for each $i = 2, \ldots, I$. Note that at time step N, each equation depends on the V for time step $N + 1$, so for each log stock price step i there is a single equation in a single unknown V_i^N. This is trivial to solve for. Given the value at time step N, one proceeds to time step $N - 1$, and again solves $I - 1$ equations, each in one unknown. Then one proceeds to time step $N - 2$, and continues in this fashion until the first time step, corresponding to the valuation date, is reached.

At each time step, the values for the claim at the lowest ($i = 1$) and highest ($i = I + 1$) stock price steps are determined from boundary conditions corresponding to the claim being valued. For an American put, for instance, $V_i^j = K - exp(\underline{Z})$ and $V_{I+1}^j = 0$. Alternatively, one can use boundary conditions that specify the first or second partial derivative at the boundary.

Early exercise can be incorporated into this process by comparing the solution to Equation 30.3 at each i with the proceeds from exercise and setting the option value equal to the larger of this solution or exercise proceeds.

The implicit method is similar but involves an additional complexity. Substituting the implicit partial derivative approximations into Equation 30.2 and rearranging produces:

$$V_i^{j+1} = A_I V_{i+1}^j + B_I V_i^j + C_I V_{i-1}^j \tag{30.4}$$

with

$$A_I = -\frac{.5\sigma^2\delta t}{\delta Z^2} - \frac{.5(r - \delta - .5\sigma^2)\delta t}{\delta Z}$$
$$B_I = \left(1 + r\delta t + \frac{\sigma^2\delta t}{\delta Z^2}\right)$$

and

$$C_I = -\frac{.5\sigma^2\delta t}{\delta Z^2} + \frac{.5(r - \delta - .5\sigma^2)\delta t}{\delta Z}$$

There is an equation for each $i = 2, \ldots, I$.

Again, valuation starts with the known values at expiration and proceeds to time step N. The value on the left-hand side of Equation 30.4 is V_i^{N+1}, which is known as of time step N. Note that in contrast to the explicit method, however, in the implicit equation (30.3), the values of the option at log price step $i-1$, i, and $i+1$ at the current time step j are present in the same equation, and all are unknown. Similarly, there is another equation that includes the values of the option at price step $i, i+1$, and $i+2$. Thus, unlike in the explicit method, in the implicit method, it is necessary to solve a set of simultaneous equations at each time step.

Moreover, additional information is needed to solve this system. Consider the equation corresponding to log stock price step I. This depends on V_{I+1}^j, which must be specified exogenously (since the equation for V_{I+1}^j would depend on V_{I+2}^j, which is outside the valuation grid). Similarly, the equation corresponding to stock price step 2 depends on V_1^j, which must also be so specified. As with the explicit method, these values at the upper and lower boundaries come from the boundary conditions appropriate for the instrument being valued.

Given the boundary conditions, one can solve the system of linear equations at time step N. In matrix notation, the set of equations can be expressed as:

$$\mathbf{V}^{\mathrm{j}} = M\mathbf{V}^{\mathrm{j}+1}$$

where $\mathbf{V}^{\mathrm{j}}$ = a $I \times 1$ vector of values of the option at each of the interior stock price points in the grid at time step j

Due to the structure of the problem, the matrix M (which contains the A, B, and C coefficients) is sparse and block diagonal. This structure facilitates efficient computational solution of the equation system. For European options, standard solution methods, such as *LU* decomposition or successive overrelaxation, can be used. For American options, when solving the system of equations, it is necessary to utilize projected successive overrelaxation to take the possibility of early exercise into account.

Crank-Nicolson also requires solution of a system of simultaneous equations.

With all three methods, one "time steps" from expiration to the valuation date, by solving the relevant equations at each time step, using values from the later in time (closer to expiration). Upon reaching the first time in the grid, the modeler will have values for each node in the time-log stock price grid.

Given these values, it is straightforward to solve for the hedging parameters (i.e., the Greeks). For instance, at time step j and log stock price step i, the option delta is:

$$\frac{\partial V}{\partial S} = \frac{1}{S}\frac{\partial V}{\partial Z} = \frac{1}{e^{Z_i}}\frac{V_{i+i}^j - V_{i-1}^j}{2\delta Z}$$

A related approximation for the option gamma is also readily obtained.

The explicit method is simple to code and easy to use because it does not require solution of a system of simultaneous equations, but this ease comes at a cost: The explicit method is not unconditionally stable, meaning that if one chooses too large a δt, the method will produce numerical garbage. The intuition behind this is straightforward. Note that the explicit method does not actually approximate a true partial derivative, because the partial derivative at time step i is approximated using values of the claim at time step $i + 1$, meaning that time is not held constant when approximating the partial derivative. This introduces inaccuracy. This inaccuracy is more severe, the bigger the time step, and can cause the values to go haywire if δt is too big. Thus, for a given choice of δZ, it is necessary to impose constraints on the size of δt in order to ensure stability.

The implicit method, in contrast, is unconditionally stable. That is, one can choose any δt without causing the method to blow up.

The explicit method is accurate to order $[\delta t, \delta Z^2]$. The implicit method is also accurate to order $[\delta t, \delta Z^2]$. Crank-Nicolson is unconditionally stable and accurate to order $[\delta t^2, \delta Z^2]$. This combination of stability and greater accuracy is a reason for the popularity of Crank-Nicolson. However, this method produces spurious oscillations, especially in values around the strike price; these oscillations are quite evident when one plots the estimates of V_{ZZ} (or V_{SS}) against the stock price. It is possible to use extrapolation in the implicit method to get second-order accuracy while avoiding the spurious oscillations in Crank-Nicolson.[5]

Finite difference methods are sufficiently flexible to value a variety of "exotic" contingent claims. For instance, barrier options (such as up-and-out or down-and-in options) are readily valued using these methods. The barriers imply boundary conditions. As noted, finite difference methods require the imposition of boundary conditions in any event, so the barriers can be naturally incorporated into the finite difference method. Other exotics that involve path dependence, such as Asian options, or look-back options, also can be priced in the finite difference approach, although these typically involve somewhat more work. Specifically, Asians and look-backs both require modification of Equation 30.1 or 30.2 to incorporate an additional state variable. This increases the dimensionality of the problem, which in turn increases computational cost. Despite this complication, finite difference methods still can be an effective and efficient way of valuing these claims (as the increase in dimensionality increases the computational costs of competing methods as well.[6]

Finite difference methods are readily and efficiently implemented on personal computers using languages such as C++ or higher-level languages such as Matlab.

HIGHER-DIMENSION PROBLEMS

The value of some derivatives depend on more than one state variable. For example, an option on the minimum of two stocks has a payoff that depends on the price of each stock. As another example, in stochastic volatility models, both the stock price and the volatility are state variables. Finite difference methods can be used to value these types of claims as well.

In the case of an option with a payoff that depends on the prices of two stocks (each of which is a GBM, and neither of which pays a dividend), there is an analog to Equation 30.1:

$$\begin{aligned} rV = {} & V_t + rS_1V_1 + .5\sigma_1^2 S_1^2 V_{11} + rS_2V_2 \\ & + .5\sigma_2^2 S_2^2 V_{22} + \rho\sigma_1\sigma_2 S_1 S_2 V_{12} \end{aligned} \tag{30.5}$$

where $S_k =$ price of stock $k = 1, 2$
$\sigma_k =$ volatility of stock $k = 1, 2$, at time t
$\rho =$ correlation between the two stock returns

As in the one-dimensional case, this equation can be discretized. If one chooses $I + 1$ stock price steps for each stock, an explicit discretization scheme results in $(I - 1)^2$ equations at each time step, each in a single unknown, that are readily solved using values from subsequent time steps in the time-stepping scheme, to approximate the option values at each of the points in the (two-dimensional) stock price grid at each time step. This method, though easy to code, suffers from the same instability problems encountered in the one-dimensional explicit scheme discussed earlier. Indeed, the time-step size required to achieve stability is even more constraining in the two- (or higher-) dimension case.

A common method for solving problems like Equation 30.5 is alternating direction implicit (ADI). This method involves creating half-time steps. At integer time steps, one uses implicit approximations for the parital derivatives with respect to stock price 1 and explicit approximations for the partial derivatives with respect to stock price 2. At half-steps, this is reversed. Note that this method is computationally expensive, as it is necessary to solve $I - 1$ systems of $I - 1$ simultaneous equations at each time step. Thus, the computational cost grows geometrically with the number of dimensions.

ADI cannot handle directly a nonzero ρ, although cumbersome transformations of the stock price variables can be used to eliminate the cross-product term. However, it is often difficult to specify appropriate boundary conditions for the transformed variables. This sharply limits the utility of this method.

Modern splitting methods developed in the Soviet Union in the 1960s avoid these problems with ADI. In the splitting method, one splits Equation 30.5 into three equations and creates two additional fractional time steps for each time step in the time grid. At the integer time step, one solves a PDE involving only the time derivative and the stock price 1 partials. At the adjacent fractional time step, one uses the values derived from the value at the integer time step as the initial conditions needed to solve a PDE involving the time partial and the stock price 2 partials. These value are then used as initial conditions to solve a PDE involving a time derivative and the cross-derivative at the second fractional time step. These solutions are then used as initial values at the next integer time step to solve a PDE involving a time derivative and S_1 partials, and so on.[7]

Splitting techniques readily handle cross-partials, and implicit solvers can be used at each integer and fractional step, ensuring stability. Note, however, the additional computational burden entailed in this method, arising from (a) the necessity of solving systems of simultaneous equations at two fractional time steps in addition to doing so at each full time step, and (b) the necessity of solving $I - 1$ systems of $I - 1$ simultaneous equations at each integer and fractional time step.

Explicit, ADI, and splitting methods can be extended (in theory) to three, four, or more dimensions, but the "curse of dimensionality"—the fact that computational costs increase geometrically in the number of dimensions—limits the utility of these methods in practice to problems involving two, and perhaps three, state variables. For higher-dimension problems, Monte Carlo techniques are preferable.

THE PROS AND CONS OF FINITE DIFFERENCE METHODS

Monte Carlo and numerical integration methods are the main competitors of finite difference methods. For low-dimension problems (one or two state variables), finite difference methods offer several advantages. They are more accurate (for given computational cost) than Monte Carlo. Moreover, it is far easier to estimate hedging parameters (the Greeks) accurately using finite difference methods. Finite difference methods also permit faster and more accurate estimation of American option values than either Monte Carlo or integration approaches. Finite difference methods also can be adapted (with some work) to solve partial derivative and difference equations that arise when underlying prices exhibit jumps.

Finite difference methods also are readily applicable to the solution of inverse problems that are sometimes encountered in finance. For instance, one explanation of the volatility smile is that volatility is an (unknown) function of the stock price. Bodurtha and Jermakyan (1999) show how to use finite difference methods and inverse techniques to infer a volatility function $\sigma(S, t)$ from a set of observed option prices. As another example, in incomplete market models, the market price of risk may be a function of state variables. For instance, the market price of volatility risk in a stochastic volatility model may be a function of the instantaneous volatility. Again, given a set of derivatives prices, finite difference and inverse techniques can be used to extract an estimate of this market price of risk function. As an application of this, Pirrong and Jermakyan (2007) extract the market price of electricity load risk from the prices of electricity forward contracts. Monte Carlo and numerical integration are not readily adapted to these applications.

The curse of dimensionality is the main shortfall of finite difference methods. Thus, although these methods are arguably preferable to Monte Carlo and numerical integration techniques for low-dimension problems, Monte Carlo dominates for higher-dimension ones.

SUGGESTED FURTHER READING

There is a rich literature on finite difference methods. Wilmott, Dewynne, and Howison (1994) present a rigorous and thorough treatment of the subject, including extensive discussion of using finite difference methods to value exotic claims. Wilmott (2006) presents an encyclopedic overview of derivatives pricing that emphasizes the finite difference valuation approach. Duffy (2006) is rigorous and readable, and is especially good in its treatment of modern splitting methods.

ENDNOTES

1. For simplicity, I will discuss implementation of finite difference methods for stock options. These same techniques are applicable, however, to other underlying instruments, including currencies and futures, whose dynamics are well described by the geometric Brownian motion.
2. Even the Black-Scholes equation for a European option must be solved numerically, as the normal cumulative terms must be approximated numerically.
3. This points out another advantage of finite difference methods. Whereas integration and Monte Carlo typically generate the value of an option for a single value of the current underlying price, finite difference methods produce the option values for a range of underlying prices.
4. The widely utilized binomial model (see, e.g., Hull, 2006) is also an explicit scheme. Although the binomial model is useful as a pedagogical tool and is widely used in practice, it faces even more acute stability problems than the trinomial method, and hence is inferior to this method and to the implicit and Crank-Nicolson methods.
5. See Duffy (2006) for details.
6. See Wilmott, Howison, and Dewynne (1994) for a detailed discussion of the application of finite difference methods to the valuation of exotic options.
7. See Pirrong (2007) for an application of splitting methods.

REFERENCES

Bodurtha, J., and M. Jermakyan. 1999. "Nonparametric Estimation of an Implied Volatility Surface," *Journal of Computational Finance* 2: 39–46.

Duffy, D. 2006. *Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach*. Hoboken, NJ: John Wiley & Sons.

Hull, J. 2006. *Options, Futures, and Other Derivatives*. New York: Prentice-Hall.

Pirrong, C. 2007. "Pricing Power Derivatives: Theory and Matlab Implementation," in J. London, ed., *Modelling Derivatives Applications in C++, Matlab, and Excel*. London: Financial Times Press.

Pirrong, C., and M. Jermakyan. 2007. "The Price of Power." Working paper, University of Houston.

Wilmott, P. 2006. *Paul Wilmott on Quantitative Finance*. Hoboken, NJ: John Wiley & Sons.

Wilmott, P., J. Dewynne, and S. Howison. 1994. *Option Pricing: Mathematical Models and Computation*. Oxford: Oxford Financial Press.

ABOUT THE AUTHOR

Craig Pirrong is professor of finance and energy markets director for the Global Energy Management Institute at the Bauer College of Business of the University of Houston. He was previously Watson Family Professor of Commodity and Financial Risk Management at Oklahoma State University and a faculty member at the University of Michigan, the University of Chicago, and Washington University. Professor Pirrong's research focuses on the organization of financial exchanges, derivatives clearing, competition between exchanges, commodity markets, the relation between market fundamentals and commodity price dynamics, and the implications of this relation for the pricing of commodity derivatives. He has published 30 articles in professional publications, is the author of three books, and has consulted widely, primarily on commodity and market manipulation–related issues. He holds a BA in economics, an MBA, and a PhD in business economics from the University of Chicago.

CHAPTER 31

Stochastic Processes and Models

GEORGE CHALAMANDARIS
Lecturer in Finance, Athens University of Economics and Business

A. G. MALLIARIS
Walter F. Mullady Sr. Professor of Business Administration, School of Business Administration, Loyola University Chicago

INTRODUCTION

Stochastic processes describe the time series behavior of random variables. Asset prices and financial concepts such as capital, debt, leverage, liquidity, dividends, profits, losses, and numerous others are random variables. If one is interested in their behavior over time, these variables are described as stochastic processes.

The simplest stochastic process is the random walk, where the variable next period is the value of the variable currently plus a random shock. Much before economists such as Black, Scholes (1973) and Merton (1973) formalized the idea of a continuous random walk in finance, the nineteenth-century British botanist Robert Brown had discovered this stochastic process by observing the random movement of pollen particles in water and suggested a mathematical description of this motion.

When a financial analyst begins with certain stochastic processes and then derives more complex relationships that are themselves stochastic processes, the branch of mathematics that describes this field is called stochastic calculus. For example, if an asset's price is described by a stochastic process and one is interested in the value of a call option on this asset, the mathematical calculations that need to be performed belong to the field of stochastic calculus.

Stochastic calculus is a relatively new branch of mathematics that deals more specifically with the calculus of functions of random processes and which was largely based on the work of Norbert Wiener (1923) and Kiyosi Itô (1951). It was motivated by engineering problems of signaling and noise. Since signaling in finance often is associated with fundamentals or trends, while noise represents randomness, the mathematical methods developed in electrical engineering found applicability in finance. Today, the ideas of stochastic calculus have become the traders' jargon in London and New York after their successful application in option pricing during the late 1970s.

The purpose of this chapter is to present some of the foundational models and properties of these probabilistic methods in order to provide the reader with the intuition behind their appropriateness and applicability in finance. A more detailed and rigorous exposition of the same ideas that are presented here can be found in Chalamandaris and Malliaris (2008), Malliaris and Brock (1982), Duffie (1996), and Karatzas and Shreve (1998).

STOCHASTIC PROCESSES

Definitions and Properties

A stochastic process is (for the purpose of this chapter) a real variable that changes through time in a random way, which is why it is usually written as $S(t,\omega) : [0,\infty) \times \Omega \to R$. The first argument describes its time dependence whereas the second argument ω denotes its dependence on the occurrence of a random event $\omega \in \Omega$ taking values out of a random set Ω.

Stochastic processes can also be characterized as discrete time processes if they are allowed to vary only at fixed time intervals or as continuous time processes if they are allowed to vary continuously in time.

Intuitively, one can look at a stochastic process as a two-dimensional entity: What we do witness in real world is a single "realized" path of the process: Starting at time $t = 0$ from value $S(0)$, we are reaching at time t its value $S(t)$, having in the meantime passed from all its intermediate realized values that together form the sample path. Mathematically we can write this as $\omega : [0,t] \to [S(0), S(t)]$. This single "realized" path is what happened for a given occurrence of ω. Now, if the unanticipated events had occurred differently (i.e., for a different $\omega' \in \Omega$), things would have evolved in another way, and we would have witnessed an alternate realized path $\omega' : [0,t] \to [S(0), S'(t)]$ with the same initial value for our process but almost certainly different intermediate and final values corresponding to each time instant. Therefore, the second dimension of a stochastic process is the chance of reaching any single path out of infinite possibilities. Exhibit 31.1 displays the dual dimensionality of a stochastic process.

The reason why most of financial quantities, such as asset prices, asset returns, and other macroeconomic factors, can be described in stochastic process formalism is almost obvious: All these quantities change through time, and they do change in an uncertain way. This uncertainty summarizes in reality the dependence of the variable in question on a large number of factors that are unobservable, too many in number, or related to it in ways too complicated to describe in a deterministic way. This is why we simply describe every influence that is too complex to anticipate as simple randomness.

For example, a stock price depends on the present value of its expected earnings. These in turn depend on uncertainty and news regarding the fundamentals of the company, on the existence of news regarding this stock, on the state of the economy and the way that investors perceive that all the above affect its prospects. Thus, the pricing of a stock quickly reaches a level of complexity that could not be possibly captured by any deterministic model.

Thus far, we have given an intuitive definition of a stochastic process. To increase the reader's knowledge about stochastic processes, we offer two very important properties that are encountered in financial modeling, namely the Markov and Martingale property.

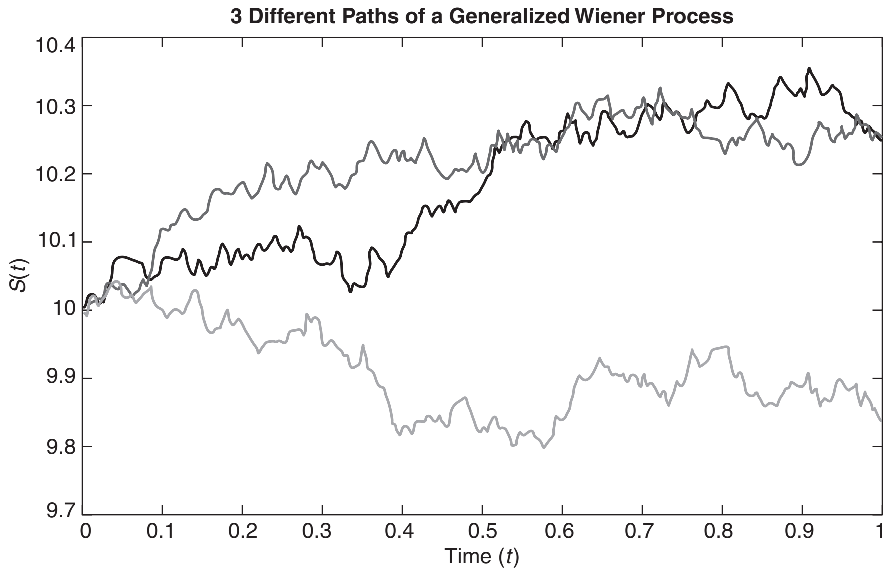


Exhibit 31.1 Three Different Realizations of a Single Stochastic Process

A *Markov process* is a stochastic process for which the current value is the only relevant one for predicting the future. A Markov process is one without "memory," since its conditional distribution for the future depends only on its current value and not on the particular path that this process followed in order to reach the present state. Mathematically we can express this property as

$$f_{S(t+1)|S(t),S((t-1),\ldots,S(0))}(S(t+1)|S(t),\, S(t-1),\, \ldots,\, S(0)) = f_{S(t+1)|S(t)}(S(t+1)|S(t)) \tag{31.1}$$

where $\quad f =$ conditional distribution function in a discrete time setting

What changes from the discrete to the continuous form is just the values that time t are allowed to take. For example in a continuous time, setting all the values $S\,(t)$ that correspond to the entire continuum between $[0, t)$ should be included in the conditional part of the distribution, suggesting the given realization path of the process until time t. A given history of the process until time t is denoted in relevant bibliography as the filtration F_t, and it implies the accumulation of all observable information up to time t. Hence, we should write

$$f_{S(t+1)|F_t}\,(S\,(t+s)\,|F_t) = f_{S(t+s)|S(t)}\,(S\,(t+s)\,|S\,(t))$$

This Markov property is very much consistent with the weak form of market efficiency, which states that all available information in a record of past prices is impounded simply in the present price of the asset.[1] Therefore, a technical trader who looks at charts of past prices in order to predict future price movements and trade accordingly cannot really expect to make above-average returns on an asset that follows a Markov process. More intuitively, the Markov property describes that the past and the future are independent; what really matters is the present, because the next value will be formed from the present value and a random shock. The past is irrelevant.

A stochastic process is a *Martingale* if, after collecting all the available information up to time t, the best forecast of the value of the process at any future time t' is no other than its current value. Technically, this can be written in terms of its expectation conditional on its current value:

$$E\left[S\left(t'\right)|F_t\right] = S(t) \tag{31.2}$$

The financial intuition behind the Martingale property will become apparent later in this chapter. It describes the idea that information is valuable and therefore it is not wasted. To put it differently, the Martingale property says that prices reflect all publically available information.

Constructing the Continuous Time Model: Brownian Motion

The elementary building block of stochastic calculus is the so-called Wiener process or standard Brownian motion. This process $B\,(t,\omega)$ satisfies these properties:

- By convention, its initial value is equal to zero almost certainly; that is, $B\,(0) = 0$.
- For any time $s > t$, the increment $B(s) - B(t)$ is normally distributed with mean zero and variance equal to $s - t$. That is,

$$E[B(s) - B(t)] = 0$$
$$Var[B(s) - B(t)] = s - t$$

- Any increments that are not overlapping are independent (i.e. for any times $0 < t_0 < t_1 < \cdots < t_n < \infty$), the random variables $B(t_0) - B(0)$, $B(t_1) - B(t_0), \ldots, B(t_n) - B(t_{n-1})$ are all independently distributed.
- For each $\omega \in \Omega$, the sample path $[0, t] \to [0, B\,(\omega, t)]$ is continuous. In other words, however closely we look at any realized path of a Brownian motion, we will never witness any discontinuities or jumps.

In words, a Brownian motion is a random process with continuous sample paths and time increments that are independent of each other and are normally distributed.

Other very important properties stem from this definition:

- The process is finite and persistent; that is, if we look at its realized paths from an increasingly closer distance, they will never look smooth however close we may get: The paths will always look jagged and random. If we look at the same paths from an increasingly greater distance, however, these paths will remain finite in the near term. To put it differently, no matter how much they fluctuate, they will not explode in short period of time.
- This property follows from the fact that the variance of its increments is proportional to their time length. Indeed, the magnitude of variability of

any random variable is comparable with its standard deviation. This means that when Δt becomes very small, the magnitude of variability of the Wiener process increments does not shrink as fast as the time interval itself, since it is scaled by $\sqrt{\Delta t}$. Similarly, when Δt becomes very large, the magnitude of variability of the random increments increases at a much slower rate just because of the square root rule. It can be proven[2] that any other scaling scheme would produce either an exploding or a decaying randomness.

- Technically, this property ensures that a Wiener process needs infinite time to reach infinity in value, and only a zero time interval would "freeze" it to no motion at all. Equivalently, it guarantees that at some finite time (however far in the future this may be) the process will (almost certainly) reach any specific target that is bounded.
- It is Markov, since the conditional distribution of any future value of the process depends only on the current value (which is a given in a conditional distribution function) and the distribution of the increment itself. We do know that this distribution is normal with zero mean and variance equal to its time length, and we also know that it is independent from any past value or increments; that is,

$$\begin{aligned} E[B(t+s)|F_t] &= E[B(t+s)|B(t)] = B(t) \\ Var[B(t+s)|F_t] &= Var[B(t+s)|B(t)] = s-t \end{aligned} \tag{31.3}$$

- It is Martingale, since its expectation conditional to the current value equals the current value itself. This follows from the fact that the expected value of any increment is zero; that is,

$$E[B(t+s)|F_t] = B(t) + E[B(s) - B(t)] = B(t) \tag{31.4}$$

- The Brownian motion is expected to hit any finite value infinitely often.
- Finally, an intriguing property of the Brownian motion is that the expected length of the path that a Brownian motion will follow within any time interval is infinite. This property will appear to be quite an obstacle when we try to employ ordinary calculus methods to deal with this process.

The Ito Process and the Need for Stochastic Calculus

These "nice" properties of the Brownian motion we just described make it an excellent candidate to become the engine of randomness in most financial models. The main idea is that any model process constructed in this manner can be structured so as to inherit these properties. A typical example of such a stochastic process with widespread application in asset price modeling is the Ito process. Algebraically it can be expressed in the form of a stochastic differential equation

$$dS(t,\omega) = \mu(S(t,\omega),t) \times dt + \sigma(S(t,\omega),\ t) \times dB(t,\omega) \tag{31.5}$$

where $\mu\,(S,t)$ = expected drift rate of the process
$\sigma\,(S,t)$ = expected variance rate (which we discuss a bit later)

Since time takes values continuously in $[0, \infty)$, the Ito equation is a continuous-time random equation. One can immediately notice that although the generator of randomness is the Brownian motion increment $dB(t, \omega)$, both the expected drift and variance rate remain stochastic through their dependence on $S(t, \omega)$, and thus randomness pervades almost all the terms in Equation 31.5. This feature might appear rather challenging when we seek for an intuitive understanding of this equation. However, this apparent complexity becomes easily resolved if we attempt a small approximation of Equation 31.5 through a discrete time setting. Indeed, if we consider to approximate dt with a very small time interval Δt, we can write Equation 31.5 as

$$\Delta S(t, \omega) = \mu(S(t, \omega), t) \times \Delta t + \sigma(S(t, \omega), t) \times \sqrt{\Delta t} \times \varepsilon \tag{31.6}$$

where $\varepsilon =$ random number drawn from a standard normal distribution

Now it becomes apparent that Equation 31.6 is equivalent to Equation 31.5 only if both $\mu(S, t)$ and $\sigma(S, t)$ remain constant for this small interval. The fact that both these variables are known at the very beginning of this time period leaves the Brownian motion increment as the only source of randomness in Equation 31.6. This feature characterizes $\mu(S, t)$ and $\sigma(S, t)$ as adapted processes to the filtration generated by the Brownian motion.

This very technical language simply states that the update of information and value in any such adapted process is strictly caused from the Brownian motion itself and specifically from its next increment. Mathematically this is equivalent to the statement that a process that is adapted to a filtration generated by a Brownian motion can be completely determined by its past trajectory leading to its current value.

It is worth paying a bit more attention to the structure of an Ito process and its financial implications. To start with, the change in price of any asset that follows an Ito process is made up by the sum of two terms. The first term $\mu(S(t, \omega), t)$ is known as the expected drift rate or in general as the drift component of the process and is meant to encapsulate the instantaneous expected value of the change in the random variable $S(t, \omega)$. Although the drift component clearly plays the role of a trend, there is no reason whatsoever to assume that this trend is in general deterministic. On the contrary, it is affected both by time and by randomness, which enters through $S(t, \omega)$ itself, spanning this way all the time and all the states of the world that are captured by our model. This indicates that when we are located in one particular state of economy at a specific time, both described by (t, ω), then we can calculate the instantaneous expected growth for our asset by $E[\mu(S(t, \omega), t)]$.

The second term $\sigma(S(t, \omega), t) \times dB(t, \omega)$ is itself the product of two factors. The first factor $\sigma(S(t, \omega), t)$ is called the diffusion coefficient or the expected variance rate and it measures the magnitude of variability of $dS(t, \omega)$ at a particular state of the world at a particular instance in time. It is in fact the scale of randomness that drives the change $dS(t, \omega)$. The second factor—the Brownian motion increment $dB(t, \omega)$—is a scale-free/normalized realization of randomness. It is simply a draw from a normal distribution that is amplified by the aforementioned diffusion coefficient.

Having already defined the diffusion coefficient as a random variable, which is adapted to the filtration generated by the Brownian motion process and therefore can be considered as constant for the infinitesimal time interval dt, it is very easy to check that

$$E[\sigma(t, S(t,\omega))dB(t,\omega)] = \sigma(t, S(t,\omega)) \times E[dB(t,\omega)] = 0 \tag{31.7}$$

$$\begin{aligned} Var[\sigma(t, S(t,\omega))dB(t,\omega)] &= \sigma^2(t, S(t,\omega)) \times Var[dB(t,\omega)] \\ &= \sigma^2(t, S(t,\omega)) \times dt \end{aligned} \tag{31.8}$$

Equations 31.7 and 31.8 are used in option pricing to calculate the risk of a portfolio. Now if we consider dt as the next trading interval, we can observe that the change in the price of an asset can be decomposed into two nonoverlapping components: the *expected* change and the *unexpected* change. The first coincides with the instantaneous growth rate, which is the anticipated variation the time/state in which we are positioned. The second component encapsulates total instantaneous uncertainty: new information that enters the model in the form of realized randomness $dB(t,\omega)$, which is next weighted by an uncertainty coefficient: the expected variance rate.

The case for using the Ito equation as a template for a family of financial models was very well made by Merton (1982), where he identified a number of foundational requirements that modern finance put forward and that Ito's process as a candidate model satisfied. Among others, he mentions that an Ito process

- Exhibits persistent randomness even as Δt becomes increasingly small.
- Evolves "well-behaved" uncertainty of finite means and variances that are all consistent with empirical evidence.
- Implies the presence of uncertainty at all trading periods.
- Maintains the independence of future from the past thanks to its Markov property. This is the way of expressing mathematically the concept of economic markets efficiency, that is, that knowledge of past price behavior does not allow for above-average returns in the future.

However, even if the formulation of this process is very tidy and financially appealing, it still required new tools of calculus to deal with it algebraically. Indeed, although Equation 31.5 seems quite similar to an ordinary differential equation, tempting us to employ the usual techniques of ordinary calculus to solve it, problems do seem to arise fast. For example, one could attempt to divide both legs by dt to bring it to an ordinary differential equation form:

$$\frac{dS(t,\omega)}{dt} = \mu(S(t,\omega), t) + \sigma(S(t,\omega), t) \times \frac{dB(t,\omega)}{dt} \tag{31.9}$$

Now we reach the first great obstacle: Brownian motion is nowhere differentiable, and hence the second part of the equation becomes impossible to define. The proper mathematical proof of this statement is beyond the scope of this chapter; however, its intuitive meaning becomes very clear if one remembers the persistence of randomness that characterizes the trajectory of a Wiener process. Indeed, differentiation of a function requires us to zoom in on the function in such a scale

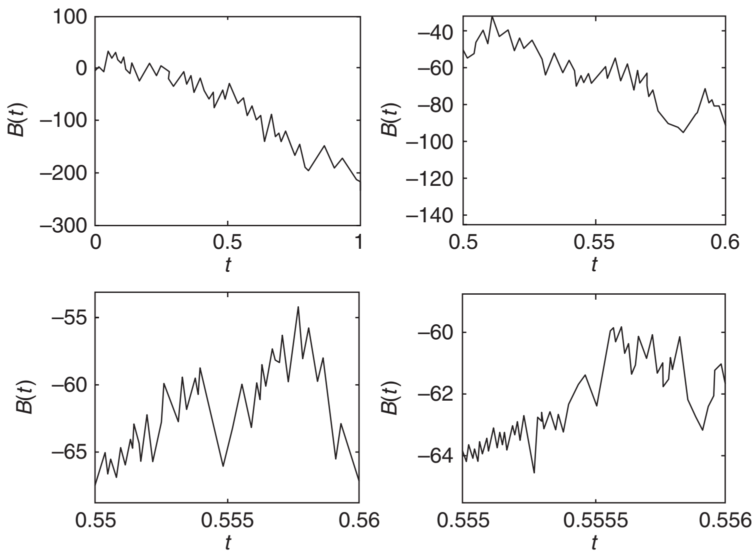


Exhibit 31.2 Zooming in on a Brownian Motion Trajectory

that the trajectory at the point of differentiation becomes a straight line. This is something that can never be done in a Brownian motion process because no matter how small we try to make the time interval we examine, the jaggedness of the path remains the same magnitude after taking into account the new scale. This implies that there exists self-similarity inside the Brownian motion as it appears in Exhibit 31.2, which makes it a fractal.

Another possibility of dealing with Equation 31.5 is to regard it as the shorthand version of an integral equation of the form

$$S(t,\omega) = S(0) + \int_0^t \mu(S(s,\omega), s) \times ds + \int_0^t \sigma(S(s,\omega), s) \times dB(s,\omega) \tag{31.10}$$

The first integral is an ordinary calculus integral (called the Riemann-Stieltjes integral) that can be computed easily. The second one, though, poses quite a difficult problem: Ordinary calculus would have us view it as a Riemann-Stieltjes integral, which we would then evaluate for each $B\,(t,\omega)$ path. However, this is an impossible task to perform in such a naive way, since all trajectories of the Brownian motion have an expected infinite length. This leaves us with only one option: to define a different kind of integral within a different kind of calculus.

BASIC ELEMENTS OF STOCHASTIC CALCULUS

Ito Integral

The definition of the Ito integral was the main breakthrough for the analytical treatment of random processes.[3] For the purposes of our exposition, we will limit ourselves to a sketch of its construction.

We write an Ito's integral as

$$I(\omega) = \int_a^b \sigma(s,\, S(t,\omega)) \times dB(s,\omega),$$

where $\sigma(s,\, S(t,\omega))$ = an integrable process that is adapted to the Brownian motion and satisfies suitable regularity and integrability conditions that are beyond the scope of this chapter
s = the time dimension of the process
ω = indicates that this path is a single occurrence of a random event

We construct this integral following the usual method: We partition the time interval $[a, b]$ into n equal intervals and we approximate the process $\sigma(t)$ with a process $\sigma_n(t)$ that is piecewise constant over each elementary time interval $[t_k, t_{k+1}]$. We can actually do that because this process is adapted at the Brownian motion $B(t)$, which simply means that its value is already known at the beginning of each $[t_k, t_{k+1}]$.

We then define an approximation of the Ito integral in a way similar to the construction of the Riemann integral, that is:

$$I_n(\omega) = \int_a^b \sigma_n(s) dB(s,\omega) = \sum_{k=0,n-1} \sigma_n(t_k) \times [B(t_{k+1},\omega) - B(t_k,\omega)] \tag{31.11}$$

One can spot immediately a major difference between Equation 31.11 and a Riemann integral: $I_n(\omega)$ cannot be anything but a random variable since it depends on the realization of the Wiener process's increments, whereas the approximating Riemann integral is always a deterministic quantity. The construction process is finally complete after driving the partition to infinite detail. It can then be proven that if the integrand satisfies certain conditions that are related to its being adapted and of finite variability, this procedure converges to a well-defined random variable:

$$I(\omega) = \int_a^b \sigma(s) dB(s,\omega) = \lim_{n\to\infty} \int_a^b \sigma_n(s) dB(s,\omega) \tag{31.12}$$

What interests us in a random variable is its distribution, and this is the case with the Ito integral. It can be demonstrated[4] that for the first two moments of its distribution, it holds

$$E[\int_0^t \sigma(s) dB(s,\omega)] = 0$$
$$\text{var}\left[\int_0^t \sigma(s) dB(s,\omega)\right] = \int_0^t E\left[\sigma(s)^2\right] ds \tag{31.13}$$

We note here that the second integral is a nonrandom, simple integral of time.

We also note that the Ito integral $\int_0^t \sigma(s)dB(s,\omega)$ is a random variable that has the property that we need only to know the history of events up to time t in order to calculate the realization of this random variable.

Ito's Lemma

In the previous section, we defined a stochastic integral, and we thus managed to take a good look at what a stochastic integral equation like Equation 31.10 stands for: It describes not a relationship of quantities that follow known or predetermined trajectories but a relationship between random variables and their dynamics. This means that Equation 31.10 defines only the distributional properties of $S(t,\omega)$ as a function of the distribution of the other random variables, with the clear implication that all equations of this kind hold "with probability 1" (or "almost certainly").

Having said that, we expect that there exists a similar solution to the problem of differentiation, since we demonstrated earlier that a Brownian motion and hence an Ito process is nowhere differentiable "with probability 1."

The problem of stochastic differentiation is posed like this: How can we differentiate a stochastic process $Y(t,\omega)$ that is a deterministic function of other Ito's processes according to $Y(t,\omega) = g(t, X(t,\omega))$?

The answer is consistent with a Taylor's expansion but with one major difference that we will now see.

Ito's Equation

Assume that $X(t,\omega)$ is an Ito process, satisfying $dX(t) = \mu(t) \times dt + \sigma(t) \times dB(t)$, and $g()$ is a deterministic, twice continuously differentiable function. Then $Y(t,\omega) = g(X(t,\omega))$ is also an Ito process, given by the stochastic differential equation

$$dY(t) = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial X} \times \mu(t) + \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times \sigma(t)^2\right) \times dt + \frac{\partial g}{\partial X} \times \sigma(t) \times dB(t) \tag{31.14}$$

A heuristic derivation of Equation 31.14 is given in the appendix at the end of this chapter. Financial analysts can use Ito's lemma when they need to derive the differential of a stochastic equation. To achieve their goal, we summarize the procedure of obtaining a differential in the result below in the more general form of Taylor's expansion. The result is known as Ito's lemma (or formula) for the univariate case:

Ito's Lemma

With assumptions as in theorem 31.14, the stochastic differential is given by

$$dY(t) = \frac{\partial g}{\partial t} \times dt + \frac{\partial g}{\partial X} \times dX + \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times dX^2$$

in which we use the formal multiplication rules

$$\begin{aligned} dt \times dt &= 0 \\ dt \times dB(t) &= 0 \\ dB(t) \times dB(t) &= dt \end{aligned} \tag{31.15}$$

The application of Ito's formula to multivariate functions of more than one Ito processes follows naturally from the last one.

Ito's Formula

Let us assume N Ito processes X_i with $i = 1, \ldots, N$ that are driven by m independent Wiener processes

$$dX_i\,(t) = \mu_i\,(X_i, t) \times dt + \sum_{k=1,\ldots,m} \sigma_{ik}\,(X_i, t) \times dB_k\,(t, w) \tag{31.16}$$

and a multivariate, deterministic, twice continuously differentiable function $g(\,)$. Then $Y = f\,(t, X_1, X_2, \ldots, X_N)$ is an Ito process with stochastic differential given by

$$dY\,(t) = \frac{\partial f}{\partial t} \times dt + \sum_{i=1:N} \frac{\partial f}{\partial X_i} \times dX_i + \frac{1}{2} \sum_{i,j=1:N} \frac{\partial^2 f}{\partial X_i \partial X_j} \times dX_i \times dX'_j \tag{31.17}$$

on which we apply the extended set of multiplication rules

$$\begin{aligned} dt \times dt &= 0 \\ dt \times dB_i(t) &= 0 \\ dB_i(t) \times dB_i(t) &= dt \\ dB_i(t) \times dB_j(t) &= 0, i \neq j \end{aligned} \tag{31.18}$$

If the driving Brownian motions are not independent but correlated, then we simply replace the last multiplication rule of Equation 31.18 with the expression $dB_i(t) \times dB_j(t) = \rho_{ij}, i \neq j$.

Example 1: Application of Ito's Lemma

As it also happens with the ordinary differential equations, there is very rarely a general solution to a stochastic differential equation. In most cases, one speculates about the form of the solution and uses the differentiation rules in order to confirm whether this is indeed a suitable one or not by trial and error. In this example, we want to confirm that

$$S\,(t) = S\,(0) \times \exp\left(\left(\mu - \frac{1}{2} \times \sigma^2\right) \times t + \sigma \times B\,(t)\right) \tag{31.19}$$

is indeed the solution to the stochastic differential equation

$$dS(t) = \mu \times S(t) \times dt + \sigma \times S(t) \times dB(t) \tag{31.20}$$

Solution
We define the Ito process

$$X(t) = (\mu - \frac{1}{2} \times \sigma^2) \times t + \sigma \times B(t)$$

with the obvious differential given by

$$dX(t) = (\mu - \frac{1}{2} \times \sigma^2) \times dt + \sigma \times dB(t)$$

We also define the deterministic, twice continuously differentiable function $g(x) = \exp(x)$. The solution we want to examine then can be written as

$$S(t) = S(0) \times \exp(X(t)).$$

The application of Ito's formula yields:

$$\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial X} = S(0) \times g(X)$$
$$\frac{\partial^2 g}{\partial X^2} = S(0) \times g(x)$$

and by substituting in Equation 31.20

$$\begin{aligned} dS(t) = {} & \left(0 + S(0) \exp(X(t)) \times \left(\mu - \frac{1}{2} \times \sigma^2\right) + \frac{1}{2} S(0) \exp(X(t)) \times \sigma^2\right) \times dt \\ & + S(0) \exp(X(t)) \times \sigma \times dB(t) \end{aligned}$$

or

$$\begin{aligned} dS(t) &= \left(0 + S(t) \times \left(\mu - \frac{1}{2} \times \sigma^2\right) + \frac{1}{2} S(t) \times \sigma^2\right) \times dt + S(t) \times \sigma \times dB(t) \\ &= dS(t) = \mu \times S(t) \times dt + \sigma(t) \times S(t) \times dB(t) \end{aligned}$$

which is the result we sought.

Equation 31.20 and its solution in Equation 31.19 are very important, because they constitute the model of the stock price in the derivation Black- Scholes formula. It is also known as a "lognormal" process because the log-changes $\ln(\frac{S(t)}{S(0)})$ follow a normal distribution with variance $\sigma^2 t$ and mean $(\mu - \frac{1}{2} \times \sigma^2) \times t$.

One can use directly Equation 31.19 to simulate realizations of $S(t)$ at time t. Furthermore, if one generates a large enough sample of realizations, this set of realizations will have distributional properties that approach asymptotically the distributional properties of Equations 31.19 and 31.20. This exactly is the idea behind Monte Carlo simulation. We demonstrate this procedure with an Excel application.

Example 2: Monte Carlo Pricing of a Call Option
Let us assume that a stock that pays no dividends follows an Ito process given by Equations 31.19 and 31.20 with an initial price $S(0) = \$18$, annualized variance rate $\sigma = 20\%$, and an annualized expected drift rate that is equal to the continuously compounded interest rate 5.00 percent, that is, $\mu = r = 5.00\%$. We want to price in Excel® with the help of a Monte Carlo simulation a call option with a 3-month expiry ($t = 0.25$) and strike $K = \$20$.

Solution
Based on the above, we can write Equation 31.19 as

$$S_i(3m) = \$18 \times \exp\left(\left(5.00\% - \frac{1}{2} \times 20\%^2\right) \times 0.25 + 20\% \times \sqrt{0.25} \times \varepsilon_i\right) \quad (31.21)$$

We note here that we have replaced the realization of the Brownian motion increment with a standardized normal random deviate ε_i. This deviate is scaled by $\sqrt{t}$ to produce the variance of the Brownian motion that corresponds to the time interval $[0, t]$.

We can now generate in Excel 10,000 random numbers that follow $N(0, 1)$ with the help of Tools-> Data Analysis-> Random Number Generator and store them in column A. We next apply Formula 31.21 on these random numbers by writing it in B2 and copy-pasting it down. Thus, the simulated stock prices at time $t = 0.25$ are stored in Column B. Finally, for each simulated realization, we calculate the realized payoff by writing the formula

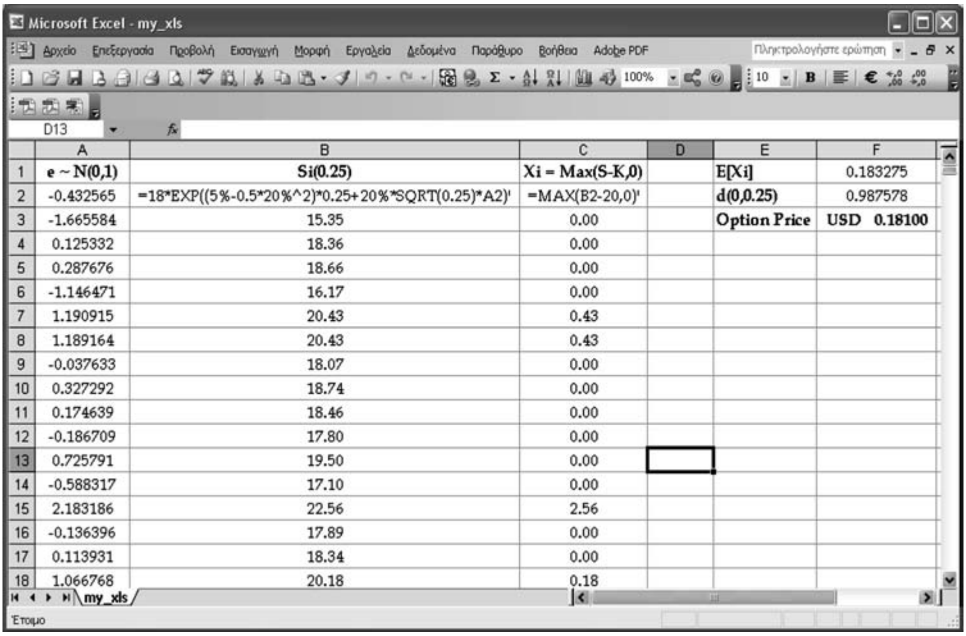
$$X_i = \max(S_i - \$20, \$0)$$

The average of these payoffs yields the expected payoff of the call option. By discounting this expected payoff using as the discount factor $d(0, 0.25) = \exp(-5\% \times 0.25)$, we get the simulated option price $c = \$0.181$. A part of the spreadsheet is displayed in Exhibit 31.3.

Calculating the same price analytically with the help of the Black-Scholes formula we get $c = \$0.1795044$, which is very close to the simulated result. By employing a larger number of simulations and using specific techniques,[5] one can further enhance the accuracy of the Monte Carlo approximation.

Example 3: Estimation of the Stock Probability Distribution
In example 2, we replicated via a Monte Carlo simulation the stock price distribution in order to price a contingent claim on it. This is the standard practice when one wants to price an exotic option that is an option with complicated or path-dependent payoff.

However, if we want to estimate only the distribution of the stock price at a future time based on Equation 31.19 or 31.20, this becomes very simple. Let us here



Microsoft Excel - my_xls

	A	B	C	D	E	F
1	e ~ N(0,1)	Si(0.25)	Xi = Max(S-K,0)		E[Xi]	0.183275
2	-0.432565	=18*EXP((5%-0.5*20%^2)*0.25+20%*SQRT(0.25)*A2)'	=MAX(B2-20,0)'		d(0,0.25)	0.987578
3	-1.665584	15.35	0.00		Option Price	USD 0.18100
4	0.125332	18.36	0.00			
5	0.287676	18.66	0.00			
6	-1.146471	16.17	0.00			
7	1.190915	20.43	0.43			
8	1.189164	20.43	0.43			
9	-0.037633	18.07	0.00			
10	0.327292	18.74	0.00			
11	0.174639	18.46	0.00			
12	-0.186709	17.80	0.00			
13	0.725791	19.50	0.00			
14	-0.588317	17.10	0.00			
15	2.183186	22.56	2.56			
16	-0.136396	17.89	0.00			
17	0.113931	18.34	0.00			
18	1.066768	20.18	0.18			

Exhibit 31.3 Monte Carlo Simulation of a Call Option Price

estimate for the settings in example 2 the probability that the stock price at $t = 1$ is larger than \$21, that is, $P(S(1) > \$21)$.

Solution

For the stock price to be larger than \$21, it takes a log-change of the stock price larger than $\ln(\frac{\$21}{\$18}) = 0.1541507$.

At $t = 1$, the log-change $\ln(\frac{S(1)}{\$18})$ follows a normal distribution with mean equal to $\mu = (5.00\% - \frac{1}{2} \times 20\%^2) \times 1 = 3.00\%$ and a standard deviation equal to $\sigma = 20.00\% \times \sqrt{1} = 20.00\%$.

Using in Excel the formula "=NORMDIST(0.1541507, 3%, 20%, TRUE)", we calculate the probability $P[\ln(\frac{S(1)}{\$18}) \leq 0.1541507] = P[S(1) \leq \$21] = 0.732619$. It follows very easily that $P[S(1) > \$21] = 1 - 0.732619 = 0.267381$.

BINOMIAL TREE: ANOTHER WAY OF VISUALIZING A STOCHASTIC PROCESS

Examples 2 and 3 demonstrated that based on the description of the stochastic process, we can identify the probability that the process will be higher or lower than a given stock price at a given time. This we can achieve analytically, either as in example 3 or with the help of a Monte Carlo procedure. In the second case, if we conduct a Monte Carlo simulation with a very large number of generated paths, then at each future time the probability that the process will pass from a certain stock price can be visualized from the density of the simulated paths that pass from its close neighborhood. In this manner, having as given a map of all possibly

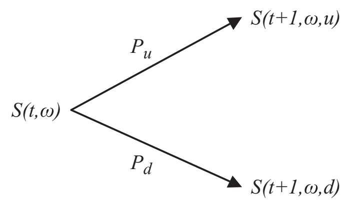


Exhibit 31.4 Multiplicative Increments

attainable future stock prices, we care only about the probability that these specific prices will be in fact attained.

Obviously, determining what is the neighborhood of a future stock price forces us to make a choice on which values are the ones that will form the centers of these neighborhoods. This is the idea behind the binomial model and all information trees that are used for pricing purposes.

Construction of a Binomial Tree and Properties

A binomial tree is a discrete time setting with discretely attainable stock prices. We consider the stock price $S(t, \omega)$ at a given time t at a given state ω. From this state, it can move to the next time step attaining only two possible state values: $S(t+1, \omega, u)$ if it goes up and $S(t+1, \omega, d)$ if it goes down, with probabilities p_u and $p_d = 1 - p_u$ respectively. Please see Exhibit 31.4.

We require that the stock process evolves through multiplicative increments so as to preclude it from reaching negative values; that is,

$$\begin{aligned} S(t+1, \omega, u) &= S(t, \omega, u) \times u, \quad u > 1 \\ S(t+1, \omega, d) &= S(t, \omega, u) \times d, \quad d < 1 \end{aligned} \tag{31.22}$$

By keeping u and d constant at all times, we make sure that we generate a tree that is recombining. This means that an "up-move" coming after a "down-move" gives the same result as a "down-move" coming after an "up-move." We start from time $t = 0$ and the single state $S(0)$ and we build incrementally the price dynamics tree by following the previous procedure. Please see Exhibit 31.5.

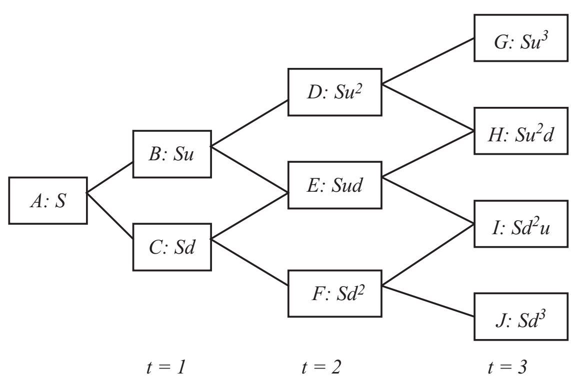


Exhibit 31.5 Markov Process

The illustrated process is a Markov process, simply because the tree is recombining. Indeed, the probability of reaching state H from state E does not depend on the path that the process followed in order to reach E at the first place. Formally,

$$P[S(t=3,\omega=H)|S(t=2,\omega=E)\wedge F_2] \\ = P[S(t=3,\omega=H)|S(t=2,\omega=E)] = p_u$$

where $F_2 =$ filtration of the stock process, the possible history of the stock process up to $t = 2$. In other words, F_2 is the set that includes all the paths that reach the every node placed at time $t = 2$: $F_2 = \{ABD, ABE, ACE, ACF\}$

Therefore, the intersection of F_2—all possible paths—with the fact that we are positioned at node E implies the only two possible paths: *A-B-E* and *A-C-E*. Mathematically we can express it as

$$S(t=2,\omega=E)\wedge F_2 = \{ABE, ACE\}$$

In the continuous time setting, we saw that the distribution of an Ito process is adequately determined by its expected drift rate and its expected variance rate. It seems very reasonable then that this should be the main information of the stock process to be conserved in a binomial tree setting as well.

Moreover, in a recombining tree, we have three free parameters to calculate—u, d, and p_u—in order to retrieve just two: drift μ and diffusion σ. This allows us a relative freedom for shaping the stock price map we were describing just earlier. The procedure of solving for the model parameters in order to retrieve a process with specific properties is known as the calibration of the model.

Cox, Ross, and Rubinstein (1979) chose $p_u = 0.5$ and solved for u and d according to

$$\begin{aligned} u &= \exp\left[\left(\mu - \frac{1}{2}\times\sigma^2\right)\times\Delta t + \sigma\times\sqrt{\Delta t}\right] \\ d &= \exp\left[\left(\mu - \frac{1}{2}\times\sigma^2\right)\times\Delta t - \sigma\times\sqrt{\Delta t}\right] \end{aligned} \tag{31.23}$$

Jarrow and Rudd (1983) chose equal jump sizes for u and d and solved thus for unequal probabilities $p_u \neq p_d$. The result was

$$\begin{aligned} u &= \exp\left[\sigma\times\sqrt{\Delta t}\right] \\ d &= \exp\left[-\sigma\times\sqrt{\Delta t}\right] \\ p_u &= \frac{1}{2} + \frac{\mu - \frac{1}{2}\sigma^2}{2\sigma}\sqrt{\Delta t} \end{aligned} \tag{31.24}$$

Other popular settings are the equal probability (EQP), and the Trigeorgis (1992) solution with optimized equal jump size.

A more insightful way to proceed is to assign the values of u and d as the ones responsible for replicating the expected variance rate and let p_u and the ensuing probability structure of the tree be responsible for the expected drift of the process.

Following Hull (2006), we treat u and d as fixed. We are looking for p_u that yields the correct expected drift of the process. Indeed, it must hold that

$$E\left[S\left(t+1,\omega\right)|S\left(t,\omega\right)\right]=S\left(t,\omega\right)\times\exp\left(\mu\times\Delta t\right)\tag{31.25}$$

By expanding the expectation we get

$$\begin{aligned}E[S(t+1,\omega)|S(t,\omega)]&=S(t+1,\omega=u)\times p_u+S(t+1,\omega=d)\times(1-p_u)\\&=S(t,\omega)\times[u\times p_u+d\times(1-p_u)]\end{aligned}$$

which, in conjunction with Equation 31.23, yields

$$\begin{aligned}&u\times p_u+d\times(1-p_u)=\exp\left(\mu\times\Delta t\right)\Rightarrow p_u=\frac{\exp\left(\mu\times\Delta t\right)-d}{u-d}\\&p_d=1-p_u\end{aligned}\tag{31.26}$$

This set of values for p_u and p_d, defines an entire probability structure for the tree: It assigns larger likelihood for reaching particular nodes and smaller one for reaching some others. This structure is also known as the probability measure of the process, and it is very interesting in that *it affects only the expected drift of the process.* For example, if we want to find the measure that makes the process a Martingale, we just need to find the value p_u that satisfies Equation 31.26 for $\mu=0$.

The resulting probability structure is known as the Martingale measure or the risk-neutral measure of the process. It is a very important measure in asset pricing. Its conceptual merit is that it is derived directly from asset prices, and it thus depends only on u and d having nothing to do with real-world probabilities. It has also been proven that if there exists a unique Martingale probability measure, then the market is complete and all contingent claims related to this stock can be priced according to it. There is a large literature on the Martingale pricing methods. Pliska (1997) and Lamberton and Lapeyre (1996) are very good references for the discrete time setting, Cochrane (2005) makes an excellent exposition of the Martingale measure in the asset-pricing framework. Chalamandaris and Malliaris (2008) provide a brief introduction to Martingale pricing methods in a continuous time setting.

Finally, we left for last the actual shaping of the stock price map: At which stock prices should we set the tree nodes? Based on the elementary probability property that the variance of a random variable Q equals $E[Q^2]-E[Q]^2$, we look for u and d values that replicate the expected variance of the stochastic process, that is:

$$\sigma^2\times\Delta t=p_u\times u^2+(1-p_u)\times d^2-[p_u\times u+(1-p_u)\times d]^2\tag{31.27}$$

By substituting Equation 31.26 in Equation 31.27 we get

$$\sigma^2 \times \Delta t = \exp(\mu \times \Delta t) \times (u + d) - u \times d - \exp(2 \times \mu \times \Delta t)$$

Ignoring all terms in Δt^2 and higher powers, we reach as solution to the equation the answer of Jarrow and Rudd in Equation 31.24 for u and d.

CONCLUSION

This chapter has introduced the reader to the definition and key properties of stochastic processes. The discussion started from the description of Brownian motion to continue to the Ito processes and to their applicability in financial modeling. It demonstrated that ordinary calculus cannot tackle the problems posed by these mathematical entities and proceeded to a brief presentation of the main concepts of stochastic calculus, namely the Ito integral and the Ito formula. Finally, the binomial tree model was discussed as an alternate way for visualizing a stochastic process and its properties.

ENDNOTES

1. See Fama (1991) for a definition of the different forms of market efficiency.
2. See, for example, Karatzas and Shreve (1998).
3. The interested reader can turn to Oksendal (1995) or to Karatzas and Shreve (1998) for a rigorous mathematical construction or to Bjork (1998) for a more heuristic derivation.
4. See Oksendal (1995) for example.
5. See, for example, Clelow and Strickland (1999).

REFERENCES

Arnold, L. 1974. *Stochastic Differential Equation: Theory and Applications*. New York: John Wiley & Sons.

Baxter, M., and A. Rennie. 1996. *Financial Calculus*. Cambridge: University Press.

Bjork, T. 1998. *Arbitrage Theory in Continuous Time*. New York: Oxford University Press.

Black, M., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, no. 3 (May/June): 637–654.

Chalamandaris, G., and A. G. Malliaris. 2008. "Ito's Calculus and the Black-Scholes Options Pricing Model," in C. F. Lee and A. Lee, eds., *Handbook of Quantitative Finance*. New York: Springer.

Clelow L., and C. Strickland. 1999. *Implementing Derivative Models*. New York: John Wiley & Sons.

Cochrane, J. H. 2005. *Asset Pricing*, rev. ed. Princeton, NJ: Princeton University Press.

Cox, J. C., S. A. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach," *Journal of Financial Economics* 7: 229–236.

Duffie, D. 1996. *Dynamic Asset Pricing Theory*, 2nd ed. Princeton, NJ: Princeton University Press.

Fama, E. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25: 383–417.

Fama, E. 1991. "Efficient Capital Markets: II," *Journal of Finance* 46: 1575–1617.

Gikhman, I., and A. V. Skorokhod. 1969. *Investment to the Theory of Random Processes*. New York: Saunders.

Hull, J. C. 2006. *Options, Futures, and Other Derivatives*, 6th ed. Upper Saddle River, NJ: Prentice-Hall.
Ito, K. 1951. *On Stochastic Differential Equations*, Memoirs of the American Mathematical Society, Number 4, Providence, Rhode Island: The American Mathematical Association Press
Ito, K., and H. McKean. 1964. *Diffusion Processes and Their Simple Paths*. New York: Academic Press.
Jarrow, R., and A. Rudd. 1982. "Approximate Option Valuation for Arbitrary Stochastic Processes," *Journal of Financial Economics* 10 (November): 347–370.
Jarrow, R., and A. Rudd. 1983. *Option Pricing*. Homewood, Illinois: R. D. Irwin.
Karatzas, I., and S. Shreve. 1998. *Methods of Mathematical Finance*. New York: Springer-Verlag.
Lamberton, D., and B. Lapeyre. 1996. *Introduction to Stochastic Calculus Applied to Finance*. London: Chapman & Hall.
Malliaris, A. G., and W. A. Brock. 1982. *Stochastic Methods in Economics and Finance*. Amsterdam: North-Holland.
Merton, R. C. 1973. "The Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4 (Spring): 141–183.
Merton, R. C. 1982. "On the Mathematics and Economics Assumption of Continuous-Time Model," in W. F. Sharpe and C. M. Cootner, eds., *Financial Economics: Essay in Honor of Paul Cootner*. Englewood Cliffs, NJ: Prentice-Hall.
Oksendal, B. 1995. *Stochastic Differential Equations*, 5th ed. New York: Springer Verlag.
Pliska, S. R. 1997. *Introduction to Mathematical Finance*. London: Blackwell.
Wiener, Norbert. 1923. "Differential space," *Journal of Mathematical Physics* 2: 131–174.
Wilmott, Paul. 2006. *Quantitative Finance*, Second Edition, New York: John Wiley & Sons.

APPENDIX: HEURISTIC DERIVATION OF ITO'S FORMULA

In this appendix we provide an intuitive derivation of Ito's lemma. For a rigorous proof, the interested reader can turn to Gikhman and Skorokhod (1969). Extensions of this lemma can be found in Arnold (1974); other heuristic derivations of the lemma can be found in Baxter and Rennie (1996), Wilmott (2006), or Bjork (1998).

Assume that the stochastic process $Y(t, \omega)$ is a deterministic function of other Ito's processes according to $Y(t, \omega) = g(t, X(t, \omega))$.

Taylor's expansion indicates that we can write $dY(t, \omega)$ as:

$$dY(t) = \frac{\partial g}{\partial t} \times dt + \frac{\partial g}{\partial X} \times dX + \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times dX^2 + \cdots \tag{31A.1}$$

In Equation 31A.1 we have suppressed ω in order to make the notation simpler and we have omitted terms of order 2 and higher for the deterministic variable t and terms of order 3 and higher for the random variable $X(t, \omega)$. We have also omitted the term $dt \times dX$ because it can be shown that it is much smaller compared to dt.

Now, since $X(t, \omega)$ is an Ito process, we can write it as

$$dX(t) = \mu(t) \times dt + \sigma(t) \times dB(t) \tag{31A.2}$$

Substituting Equation 31A.2 in Equation 31A.1 we get

$$dY(t) = \frac{\partial g}{\partial t} \times dt + \frac{\partial g}{\partial X} \times (\mu(t) \times dt + \sigma(t) \times dB(t))$$
$$+ \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times (\mu(t) \times dt + \sigma(t) \times dB(t))(\mu(t) \times dt + \sigma(t) \times \mathrm{dB}(t)) + \cdots$$

or

$$dY(t) = \frac{\partial g}{\partial t} \times dt + \frac{\partial g}{\partial X} \times \mu(t) \times dt + \frac{\partial g}{\partial X} \times \sigma(t) \times dB(t) + \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times \mu(t)^2 \times dt^2$$
$$+ \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times \sigma(t)^2 \times dB(t)^2 + \frac{\partial^2 g}{\partial X^2} \times \sigma(t) \times \mu(t) \times dB(t) \times dt \tag{31A.3}$$

Again in Equation 31A.3, dt^2 is negligible compared to dt, whereas we can again neglect $dB(t) \times dt$ for the same reason. The term that remains of interest is $() \times dB(t)^2$. Let us try to approximate the integral $\int_0^t dB\,(s)^2$ with what we know: that is, we make a very fine partition of the interval $[0, t]$ at n subintervals and we write

$$\int_0^t dB(s)^2 = \sum_{i=1}^{n}\left(B\left(\frac{t \times i}{n}\right) - B\left(\frac{t \times (i-1)}{n}\right)\right) \tag{31A.4}$$

Now if we define

$$Z_n(i) = \frac{B\left(\frac{t \times i}{n}\right) - B\left(\frac{t \times (i-1)}{n}\right)}{\sqrt{t/n}} \tag{31A.5}$$

then for each partition n we have defined a sequence $Z_n\,(1)\,,\ Z_n\,(2)\,,\ Z_n(3),\ \ldots$ that is, independent and identically distributed, and follow the standard normal distribution $N\,(0, 1)$. Using Equation 31A.5 we can rewrite our approximation of Equation 31A.4 as

$$\int_0^t \mathrm{dB}(s)^2 \approx t \times \sum_{i=1}^{n}\frac{Z_n(i)^2}{n} \tag{31A.6}$$

The second part of Equation 31A.6 converges by the weak law of large numbers toward the constant expectation of each $Z_n(i)^2$, that is, 1. Hence, we reach the conclusion that $\int_0^t dB(s)^2 = t$ or in differential form $dB(s)^2 = dt$.

Hence, substituting this result in Equation 31A.3, we get Ito's formula, Equation 31.14.

$$dY(t) = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial X} \times \mu(t) + \frac{1}{2}\frac{\partial^2 g}{\partial X^2} \times \sigma(t)^2\right) \times dt + \frac{\partial g}{\partial X} \times \sigma(t) \times dB(t)$$

ABOUT THE AUTHORS

George Chalamandaris is a lecturer in finance at the Athens University of Economics and in business at the Department of Accounting and Finance. He graduated from the National Technical University of Athens and continued his studies at Imperial College in London, where he completed his PhD in quantitative finance.

He has worked in Natwest Markets in London as well as in Emporiki Investment Bank and EFG Eurobank in Athens, holding senior positions in the areas of risk management, financial engineering, structuring, and alternative investments.

His main research interests are focused in the field of implied volatility surfaces, credit derivatives, and asset pricing involving dynamic trading strategies. He has spoken in various academic and practitioners' conferences and has published in academic journals and books in the area of derivatives pricing and risk.

A. G. Malliaris is professor of economics and finance at Loyola University Chicago and holds the Walter F. Mullady Sr. Chair in Business Administration. He graduated from the Athens School of Economics and Business with a BA in economics and earned a PhD in economics from the University of Oklahoma and a second PhD in mathematics from the University of Chicago. He has coauthored with William A. Brock two books, *Stochastic Methods in Economics and Finance* and *Differential Equations, Stability and Chaos in Dynamic Economics.* He has also authored two other books, *Foundations of Futures Markets* (1999) and *Economic Uncertainty, Instabilities and Asset Bubbles* (2005). He specializes in financial economics and has made several contributions in the area of futures and options markets.

CHAPTER 32

Measuring and Hedging Option Price Sensitivities

R. BRIAN BALYEAT
Assistant Professor of Finance, Xavier University

Like any other financial asset, the price of an option is sensitive to various risks. From the Merton (1973) model, there are six inputs affecting option prices. These are the stock price, S; the strike price, K; the risk-free rate, r; the continuous dividend yield, q; time to expiration, T; and the volatility of the stock, σ. This chapter examines option price sensitivities to these inputs and how to hedge these risks. These sensitivities are sometimes called the Greeks.

The Merton model derives the value of a put and a call on a stock as

$$
\begin{aligned}
c &= Se^{-qT}N(d_1) - Ke^{-rT}N(d_2)\\
p &= Ke^{-rT}N(-d_2) - Se^{-qT}N(-d_1)\\
d_1 &= \frac{\ln(S/K) + (r - q + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\\
d_2 &= d_1 - \sigma\sqrt{T}
\end{aligned}
$$

where $N(\bullet)$ = cumulative normal function

In the case where $q = 0$, the equation is the familiar Black-Scholes (1973) model. Sensitivities to each of the six inputs are calculated as first derivatives of the option price relative to the desired input. Because these sensitivities are derivatives of the option pricing function, they measure changes in the value of an option for small changes in the parameter.

Some of the parameters in the Merton model are more likely to change than others. Over the life of an option, the value of the stock and its volatility can change greatly. However, the risk-free rate and the dividend yield are less likely to change dramatically. While the strike price of an option generally does not change, the time to maturity of an option changes or decays in a predictable manner. The most important option price sensitivity is to changes in the stock price.

DELTA

Delta, Δ, is the rate of change of the option price relative to changes in the stock price. As a first derivative, delta measures the slope of the tangent at a given stock

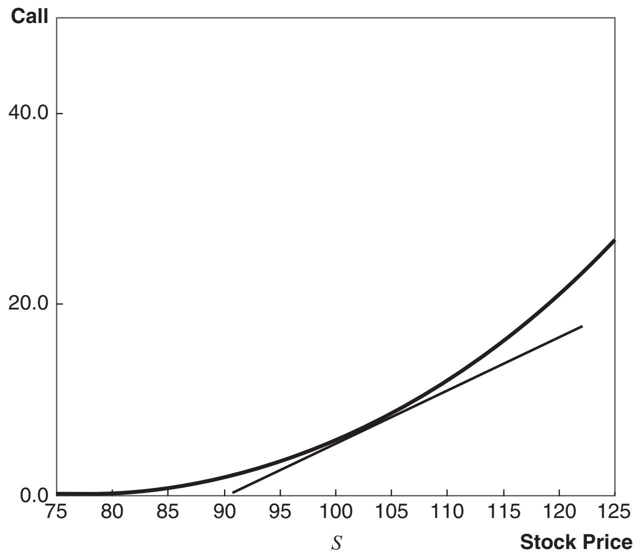


Exhibit 32.1 Call Values Prior to Maturity

price. Exhibit 32.1 graphs the value of a call option as a function of the stock price for the base case parameters presented in Exhibit 32.2. The slope of the tangent line at $S = 100$ is the delta for the at-the-money option. Close to the current stock price of \$100, the tangent approximates the call fairly well. However, for large movements in the stock price, the tangent is less accurate.

For an option, delta is the first derivative of the option price with respect to the stock price. Mathematically, for a call option

$$\begin{aligned}\Delta_c &= \frac{\partial c}{\partial S} \\ &= e^{-qT} N(d_1)\end{aligned} \tag{32.1}$$

and for a put option,

$$\begin{aligned}\Delta_p &= \frac{\partial p}{\partial S} \\ &= e^{-qT} [N(d_1) - 1]\end{aligned} \tag{32.2}$$

Exhibit 32.2 Base Case Parameters

Parameter	Input
S	100
X	100
r	6%
q	1%
T	91 days
σ	25%

Because the terminal payoff for a call option is max[0, $S - K$], as the stock price rises, the call option also increases in value. Likewise, given that the terminal payoff for a put is max[0, $K - S$], as the stock price rises, the value of put options fall. Because e^{-qT} and $N(d_1)$ are both always positive and $N(d_1)$ is always less than 1, the delta for a call is always positive and the delta for a put is always negative.

For parameters in Exhibit 32.2, the delta for the call is 0.5631 and for the put is −0.4345. Thus, for small changes in the stock price, the at-the-money call option represented by the parameters in Exhibit 32.2 will change by 56 percent of the change in the stock price. Likewise, for small changes in the stock price, the corresponding put will change by 43 percent of the change in the stock price, but in the opposite direction.

Example 1

For the parameters given in Exhibit 32.2, the price of the call is \$5.58 and the price of the put is \$4.34. Given the deltas for these options computed earlier, what is the expected new put and call price if the stock increases in value by \$1?

If the stock increases by \$1, we would expect the call to increase by \$1 × 0.5631 = \$0.56. The resulting new price would the original price of \$5.58 plus the predicted change of \$0.56 for a resulting price of \$6.14. The put would change in value by −\$0.43 = \$1 × −0.4345 to \$3.91. The actual put and call prices for a stock price of \$101 are \$3.92 and \$6.15, respectively. In each case, our estimates are only off by about a penny.

Example 2

What if the stock increases to \$105?

If the stock increases by \$5, we would expect the call to increase by \$2.82 = \$5 × 0.5631 to \$8.40 and the put would change in value by −\$2.17 = \$5 × −0.4345 to \$2.17. The actual put and call prices for a stock price of \$105 are \$8.77 and \$2.54, respectively. In each case, our estimates are only off by considerably more than a penny.

As noted earlier, as a first derivative, delta is accurate only for small changes in the stock price. The larger the change in the stock price, the less accurate the approximation. The approximation becomes less accurate because as the stock price changes, the delta for both the put and the call change as well. This can easily be seen in Exhibit 32.1 for the case of the call option. As the stock price increases, the slope of the tangent and thus delta increase as well. Likewise, as the stock price decreases, the slope of the tangent and thus delta decrease as well. Exhibit 32.3 graphs delta as a function of the stock price for both puts and calls.

For deep out-of-the-money call options, delta is close to zero. Because the call is deep out-of-the-money, a small increase in the stock price will have very little impact on the option price. As the stock price increases, the delta for the call option increases. For at-the-money call options, the delta is close to 0.5. As the stock price continues to increase, the delta for the call option asymptotes toward 1. For deep in-the-money options, there is almost a one-to-one correspondence between the change in the stock price and the change in the call option.

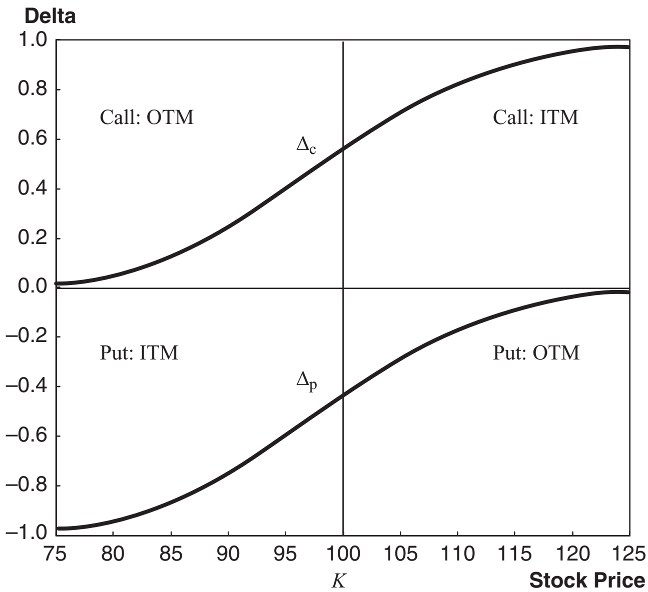


Exhibit 32.3 Delta as a Function of Stock Price

For deep in-the-money put options, the delta is close to -1. For every dollar the stock price increases, deep in-the-money put options lose almost a dollar in value. The delta for an at-the-money put option is close to -0.5. As put options get farther and farther out-of-the-money, their deltas approach zero from below. At this point, further increases in the stock price have little impact on the option value as it is very close to zero already.

Exhibit 32.4 examines put and call deltas as a function of strike price. For both puts and calls, the lower the strike price, the higher the delta. Likewise, the higher the strike price, the lower the delta.

Both Exhibits 32.3 and 32.4 illustrate the symmetry between the delta for a call and the delta for a put. The distance between the delta for a call and the delta for a put with the same strike price in Exhibit 32.4 appears to be constant. The put-call parity can be used to show that the sum of the absolute value of the delta for a call and the absolute value of the delta for the corresponding put must sum to e^{-qT}.

The intuition behind the put-call parity is that a portfolio that long a call and is short the corresponding put has an identical terminal payoff to a portfolio that is long e^{-qT} units of the underlying and is short the present value of the strike price. Thus, the portfolios must also have the same initial value. Given that the put and the call in the first portfolio have the same strike price, only one of the options can be in-the-money at expiration. If the long call is in-the-money, the payoff is $S - K$, and if the short put is in-the-money the payoff is $-(K - S) = S - K$. Thus, in either case, the terminal payoff for a portfolio that is long a call and is short the corresponding put is $S - K$ regardless of the strike price or the final stock price. Additionally, reinvesting the dividends back into the stock, a portfolio that starts with e^{-qT} units of stock will grow to one unit of stock worth S and the short position in the present value of the strike price, $-Ke^{-rT}$, will grow at the risk-free rate to

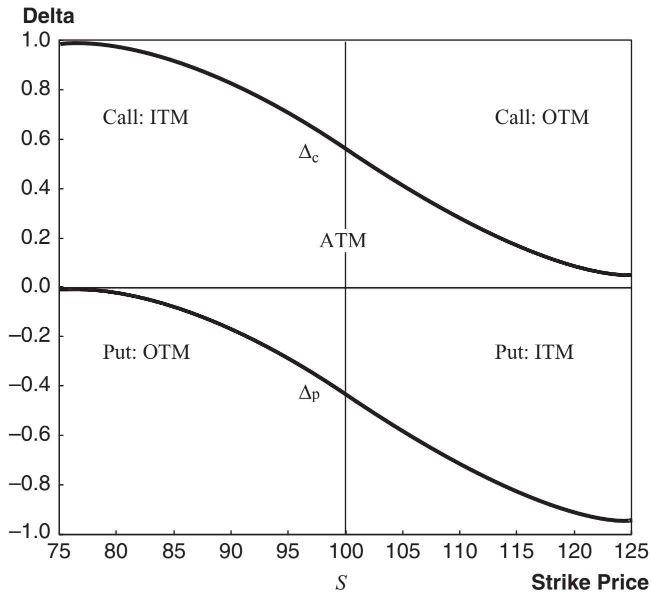


Exhibit 32.4 Delta as a Function of Strike Price

$-K$. Since both portfolios have the same terminal payoff, to avoid an arbitrage opportunity, they must be worth the same today.

Thus, from the put-call parity, we have

$$c - p = Se^{-qT} - Ke^{-rT}$$

Differentiating each side of the put-call parity with respect to the stock price and separating terms yields

$$\frac{\partial(c-p)}{\partial S} = \frac{\partial(Se^{-qT} - Ke^{-rt})}{\partial S}$$
$$\frac{\partial c}{\partial S} - \frac{\partial p}{\partial S} = \frac{\partial(Se^{-qT})}{\partial S} - \frac{\partial(Ke^{-rt})}{\partial S}$$

Noting that the derivative of the stock price with respect to itself must be 1, that the strike price of an option is not a function of the stock price, and using the definition of delta, we have

$$\Delta_c - \Delta_p = e^{-qT} - 0$$

Given that the delta for a stock is always positive and that the delta for a put is always negative yields

$$|\Delta_c| + |\Delta_p| = e^{-qT}$$

In the case where $q = 0$, $e^{-qT} = 1$ and the result reduces to

$$|\Delta_c| + |\Delta_p| = 1$$

A portfolio is considered delta-neutral if the value of the portfolio does not change when the stock price changes. The concept of delta-neutral portfolios is integral to both the Black-Scholes option pricing model and the binomial model of Cox, Ross, and Rubinstein (1979). The delta for a portfolio is the sum of the delta exposure of each asset in the portfolio. Thus, the delta for a portfolio, Δ_Π, of N different assets types is

$$\Delta_\Pi = \sum_{i=1}^{N} n_i \Delta_i$$

where n_i = number of assets of each type held with a delta of Δ_i

As noted earlier, the delta for a call option matching the parameters in Exhibit 32.2 is 0.5631. If one forms a portfolio by shorting this option, the delta is -0.5631. Because the delta for a share of stock is 1, this portfolio can be made delta-neutral by adding 0.5631 shares of stock. Exhibit 32.5 examines the performance of this portfolio for stock price changes of plus and minus 10 percent.

If the stock price remains at \$100, then the option is worth \$5.58 and the 0.5631 shares are worth \$56.31. Because the portfolio is short the call option and long the shares, the value of the portfolio is $-5.58 + \$56.31 = \$50.73$. Since the stock price is still at the base case level of \$100, the value of the portfolio remains unchanged for a return of zero percent. When the stock price rises to \$110, the call is worth \$12.60 and the 0.5631 shares are worth \$61.94 for a portfolio value of \$49.33. The value of the portfolio drops from \$50.73 to \$49.34 for a loss of \$1.39. This loss equates to a return of -2.75 percent. Likewise, when the stock price decreases by 10 percent, the value of the portfolio decreases by \$1.61 for a return of -3.17 percent.

In the binomial model, a delta-neutral portfolio is possible because at any point in the interior of the binomial lattice, there are only two possible future stock prices. In the Black-Scholes model, the portfolio is delta neutral instantaneously. In the example in Exhibit 32.5, the portfolio did not remain delta-neutral. This is because delta changes when the stock price changes. From Equation 32.1, delta for a call is a function of $N(d_1)$. When the stock price changes, d_1 changes and thus $N(d_1)$ and delta change. Exhibit 32.6 shows how a changing delta affects the hedging performance of a delta-neutral portfolio.

Exhibit 32.5 Delta-Neutral Portfolio for Short 1 Call Option

Stock Price	Call	Shares	Portfolio	Portfolio Change	Return
90	1.55	50.67	49.12	−1.61	−3.17%
100	5.58	56.31	50.73	0.00	0.00%
110	12.60	61.94	49.34	−1.39	−2.75%

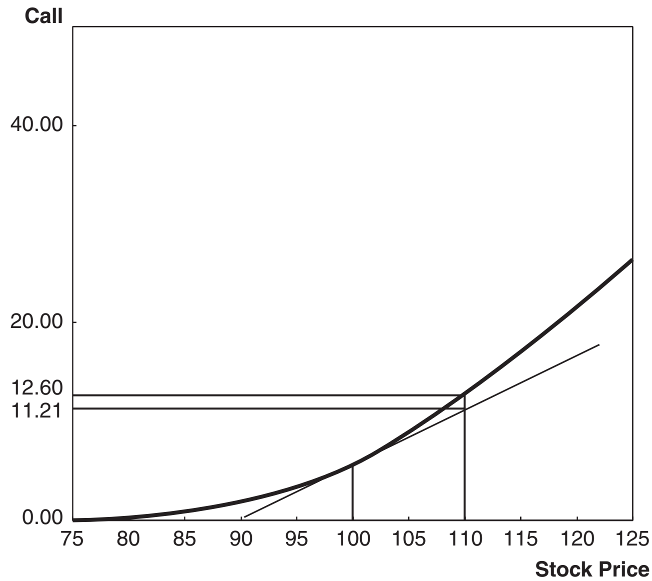


Exhibit 32.6 Delta Hedging Error

When the stock price changes from \$100 to \$110, the expected new call option value is expected to increase by $0.5631 \times \$10 = \$5.63$. This leads to a new expected call option price of $\$5.58 + \$5.63 = \$11.21$. However, due to the curvature of the option pricing function, the resulting call option price for a stock price of \$110 is \$12.60. The difference between the expected option price of \$11.21 and the resulting option price of \$12.60 is \$1.39. This is the same difference as the decrease in the value of the delta-neutral portfolio shown in Exhibit 32.5. This loss occurred because the portfolio is only delta-neutral for small changes in the underlying. As noted in Exhibit 32.3, as the stock price increases, the call option becomes in-the-money and the delta for the option increases. This effect also can be seen in Exhibit 32.6, where the call option pricing function becomes steeper as the stock price increases. Please see Exhibit 32.7.

Exhibit 32.7 examines the performance of a delta-neutral portfolio for the at-the-money put. The delta for the at-the-money put is −0.4345. Thus, a portfolio that is long one put option can be made delta-neutral by purchasing 0.4345 shares. Unlike the delta-neutral portfolio for the call, the delta-neutral portfolio for the put has a small positive return. The different return for the two examples results from

Exhibit 32.7 Delta-Neutral Portfolio for Long 1 Put Option

Stock Price	Put	Shares	Portfolio	Portfolio Change	Return
90	10.29	39.10	49.39	1.60	3.37%
100	4.34	43.45	47.79	0.00	0.00%
110	1.39	47.79	49.18	1.39	2.92%

the exposure of the portfolios to changes in delta as the stock price changes. This exposure is measured by the gamma of the portfolio.

GAMMA

Gamma, Γ, is the rate of change in the option delta relative to changes in the stock price. Thus, it is also the second derivative of the option price relative to the stock price. Mathematically, the gamma for a call option is

$$\begin{aligned}\Gamma_c &= \frac{\partial^2 c}{\partial S^2} \\ &= \frac{\partial \Delta_c}{\partial S} \\ &= \frac{e^{-qT} N'(d_1)}{S_0 \sigma \sqrt{T}}\end{aligned} \tag{32.3}$$

where $N'(d_1)$ = first derivative of the cumulative density function for the normal distribution.

The first derivative of a cumulative density function is the probability density function. In the case of the standard normal, the probability density function is

$$N'(d_1) = n(d_1)\frac{1}{\sqrt{2T}}e^{-(d_1)^2/2}$$

The gamma for a put and the gamma for the corresponding call are identical. This result can be proven using the put-call parity. As with the example for delta, first differentiate both sides of the put-call parity with respect to the stock price and apply the definition of delta.

$$\begin{aligned}c - p &= Se^{-qT} - Ke^{-rT} \\ \frac{\partial(c-p)}{\partial S} &= \frac{\partial(Se^{-qT} - Ke^{-rT})}{\partial S} \\ \Delta_c - \Delta_p &= e^{-qT}\end{aligned}$$

Next, differentiate both sides again with respect to the stock price and separate terms.

$$\begin{aligned}\frac{\partial(\Delta_c - \Delta_p)}{\partial S} &= \frac{\partial(e^{-qT})}{\partial S} \\ \frac{\partial(\Delta_c)}{\partial S} - \frac{\partial(\Delta_p)}{\partial S} &= 0\end{aligned}$$

Applying the definition of gamma as the first derivative of delta with respect to the stock price and a little algebra yields

$$\Gamma_c - \Gamma_p = 0$$
$$\Gamma_c = \Gamma_p$$

Hence, $\Gamma_c = \Gamma_p = \Gamma$. Additionally, as in the case of delta, the gamma for a portfolio is the sum of the gamma exposure of each asset in the portfolio.

The gamma for both the put and call using the parameters in Exhibit 32.2 is 0.0315.

We can use this gamma to calculate the expected delta for changes in the stock price.

Example 3

For the parameters given in Exhibit 32.2, the delta of the call is 0.5631 and the delta for the put is −0.4345. What are the expected deltas for the put and the call when the stock price increases to $110?

The stock price rising to $110 is a $10 increase in the stock price. Given a gamma of 0.0315, a $10 increase in the stock price should increase the delta of the put and the call by $0.315 = 0.0315 \times 10$. The resulting expected delta for the call is 0.8777 and the expected delta for the put is −0.1198. The actual new deltas are 0.8207 and −0.1768 for the call and the put, respectively. While the estimates for the new delta are close, they are not exact. Thus, even gamma changes as the stock price changes. Exhibit 32.8 graphs the relationship between the stock price and gamma.

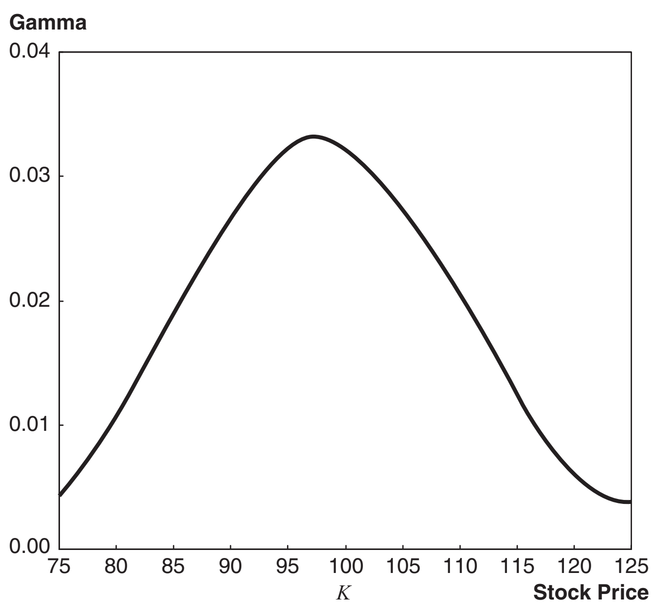


Exhibit 32.8 Gamma as a Function of Stock Price

Gamma is the largest for near- or at-the-money options and decreases the farther one moves away from this level. The smaller the gamma, the less delta changes for a given change in the underlying. The pricing function for deep in-the-money and deep out-of-the-money puts and calls is relatively flat. Thus, the deltas for these options are not highly responsive to changes in the option price. The gamma for these options is relatively small. Delta changes the most for near-the-money options. Thus, these options have the largest gamma. Additionally, gamma is always positive. This can be seen mathematically as each term in Equation 32.3 must be positive. Also, the graphs in Exhibit 32.3 for the delta of a put and call are increasing functions of the stock price. Hence, the slopes of these functions (i.e., gamma) must be positive.

The example in Exhibit 32.5 included shorting a call option and purchasing 0.5631 share of stock. Because the delta for a share of stock is always 1 and gamma measures how delta changes relative to the stock price, gamma for a share of stock is always zero. Thus, the gamma for the portfolio is equal to -1 times the gamma of the at-the-money call plus zero. Because the gamma of a long call must be positive, the gamma of the portfolio is negative. Likewise, since the example in Exhibit 32.7 involved purchasing a put option, the gamma of that portfolio is positive.

Delta-neutral portfolios with negative gamma lose money when the stock price changes. If the stock price increases, the negative gamma causes the delta of the portfolio to become negative; thus any further increases in the stock price decrease the value of the portfolio. If the stock price falls, the negative gamma causes the delta of the portfolio to become positive; thus any further decreases in the stock price decrease the value of the portfolio. Likewise, delta-neutral portfolios with positive gamma make money when the stock price changes. This explains the profit patterns present in Exhibit 32.5 and 32.7.

To hedge the option positions presented in Exhibit 32.5 and 32.7 more accurately, one would need to be both delta-neutral and gamma-neutral. Making portfolios both delta-neutral and gamma-neutral is the subject of the last section of this chapter. Making a portfolio both delta-neutral and gamma-neutral is similar to adding a convexity adjustment to duration in bond pricing applications.

Thus, gamma proxies for the curvature in the option function across stock prices. As noted earlier, this curvature is greatest for at-the-money options. The curvature decreases for both puts and calls as one moves farther away from the at-the-money level. Additionally, the curvature is a function of the moneyness of the option and the time to maturity for the option. Exhibit 32.9 graphs gamma as a function of time to maturity for at-the-money options, where $S > K$ (in-the-money calls, out-of-the-money puts), and where $S < K$ (out-of-the-money calls, in-the-money puts).

Even a portfolio of options that is delta and gamma-neutral will change in value over time. For example, even if an at-the-money put is both delta and gamma hedged, the portfolio loses value if the stock remains \$100. This is because as at-the-money options get closer to maturity, all else being equal, they decrease in value. If the stock price does not change, the at-the-money option expires worthless. This phenomenon is known as the time decay in option prices and is represented by theta.

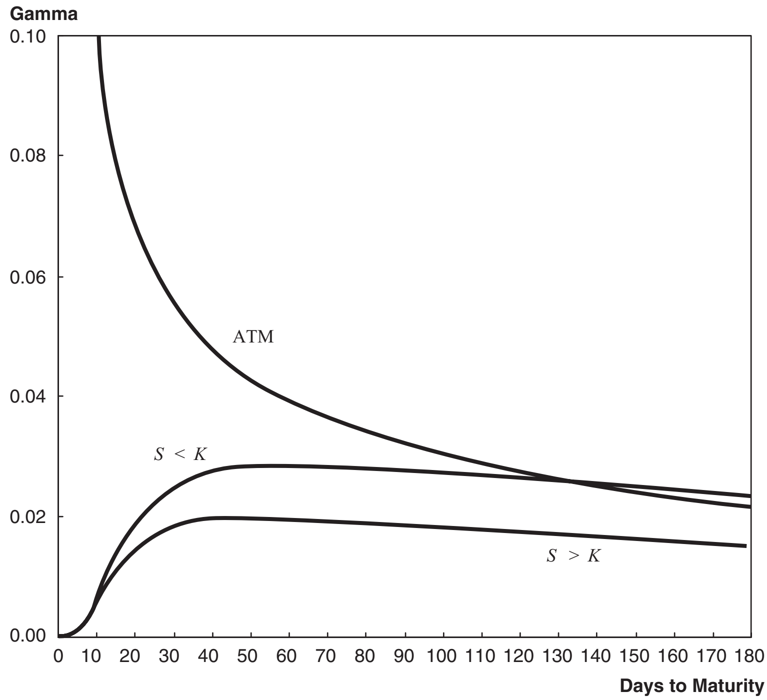


Exhibit 32.9 Gamma as a Function of Time to Maturity

THETA

Theta, Θ, is the rate of change in the option price relative to the passage of time. The time to maturity for an option decreases as an option matures. Because this change in time is measured as a positive number, theta is defined to be -1 times the first derivative of the option price relative to time. Mathematically, the theta for a call option is

$$\begin{aligned}\Theta_c &= -\frac{\partial c}{\partial T} \\ &= -\frac{Se^{-qT}N'(d_1)\sigma}{2\sqrt{T}} + q\,Se^{-qT}N(d_1) - r\,Ke^{-rT}N(d_2)\end{aligned} \tag{32.4}$$

and the theta for a put option is

$$\begin{aligned}\Theta_p &= -\frac{\partial p}{\partial T} \\ &= -\frac{Se^{-qT}N'(d_1)\sigma}{2\sqrt{T}} - q\,Se^{-qT}N(-d_1) + r\,Ke^{-rT}N(-d_2)\end{aligned} \tag{32.5}$$

The theta for the call using the parameters in Exhibit 32.2 is −12.31. The theta for the corresponding put is −7.40. Both of these thetas are computed where the passage of time is measured in years. To calculate a theta based on calendar days, simply divide the theta calculations in Equations 32.4 and 32.5 by 365. Likewise, theta can be converted into trading days by dividing by 252.

Example 4

The maturity of the options in Exhibit 32.2 is 91 days. Assuming that none of the parameters changes, what are the expected prices of the put and call the next day?

Dividing the previous theta calculations by 365 yields a daily theta of −0.0337 for the call and −0.0203 for the put. Thus, the call is expected to lose about 3 cents in value per calendar day and the put is expected to lose about 2 cents. This results in an estimated cost for the 90-day call of \$5.55 and the 90-day put of \$4.32. Computing the put and call prices using the parameters in Exhibit 32.2 for 90-day options gives almost identical results.

For most options, theta is negative. The two exceptions are deep in-the-money puts and deep out-of-the-money calls with very large dividend yields. Exhibits 32.10 and 32.11 graph the value of a put and call with the parameters in Exhibit 32.2 as a function of stock price.

For stock prices above \$90, the value of the put in Exhibit 32.10 is larger than the intrinsic value of the option. All else being equal, these options will decay in value as they approach maturity. Thus, these options have a positive theta. However, for sufficiently deep in-the-money options, the value of the put is less than the intrinsic value of the option. These options have a positive theta. All else being equal, these options will increase in value over time.

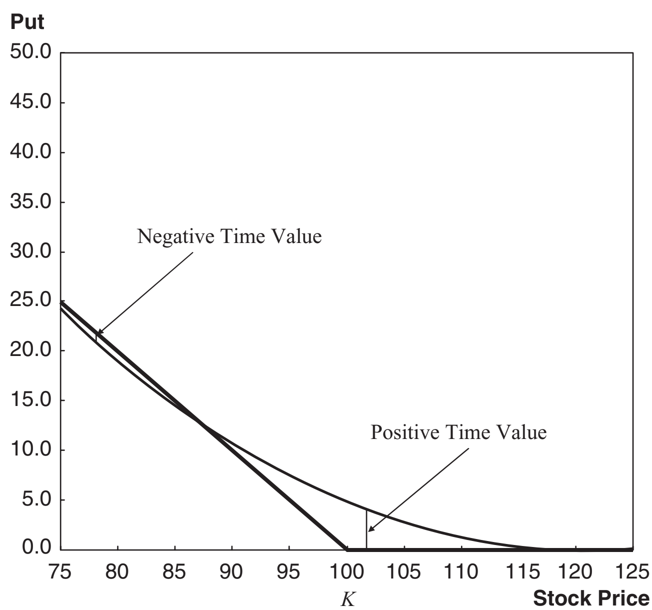


Exhibit 32.10 Put Values Prior to Expiration

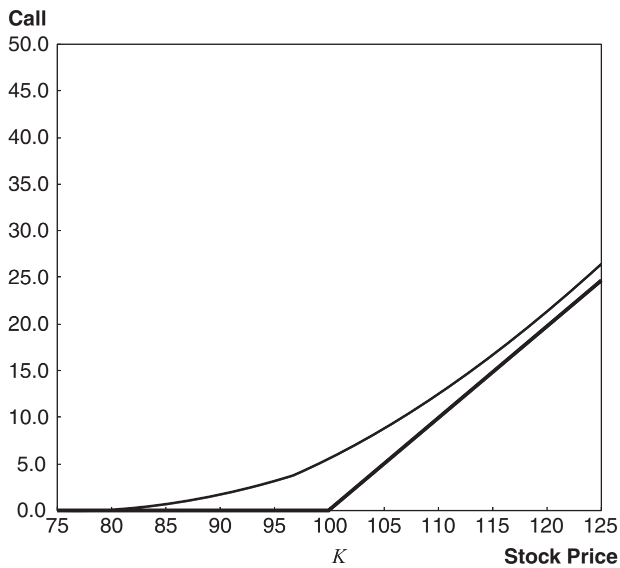


Exhibit 32.11 Call Values Prior to Expiration

As can be seen in Exhibit 32.11, for the parameters in Exhibit 32.2, the value of the call option is always greater than the intrinsic value. Hence, for the given stock prices, the theta of the call option is negative. Exhibit 32.12 graphs the theta for the put and the call as a function of strike price.

Theta is negative for the entire range except for the deep in-the-money put options. Theta is the smallest for the near at-the-money options and increases as

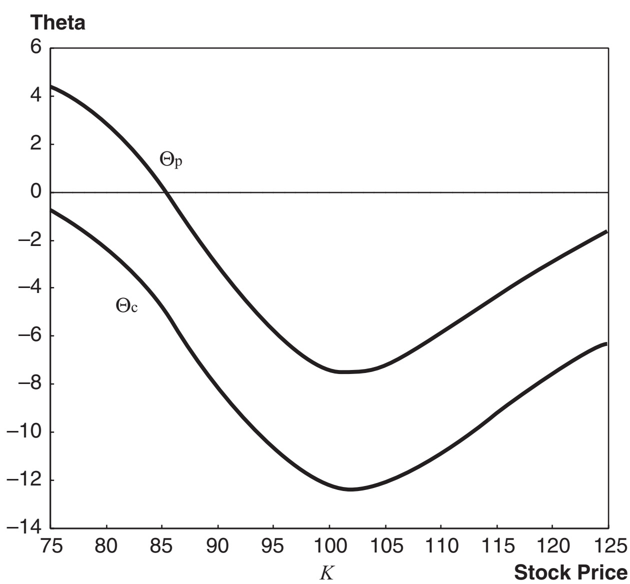


Exhibit 32.12 Theta as a Function of Stock Price

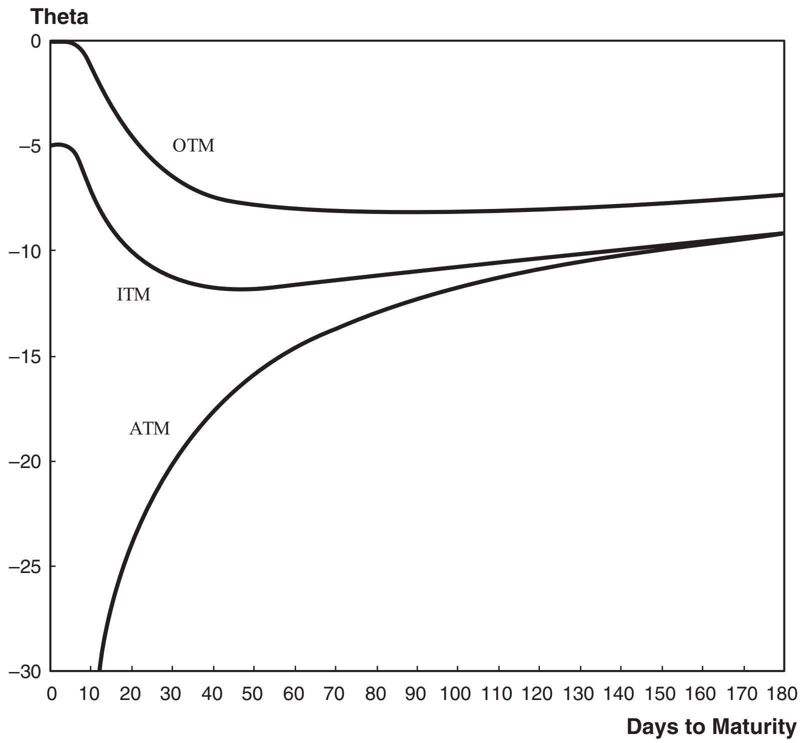


Exhibit 32.13 Call Thetas as a Function of Time to Maturity

one moves away from the at-the-money level. The pattern for the thetas across strike prices is the inverse of the pattern for gamma: Where theta is largest, gamma is small, and vice versa. Exhibits 32.13 and 32.14 graph the relationship between theta and time to maturity for in-the-money, at-the-money, and out-of-the-money puts and calls.

The pattern of the thetas across time is also inversely related to gamma across time. The relationship among delta, gamma, and theta can be represented by the Black-Scholes-Merton differential equation

$$\frac{\partial \Pi}{\partial T} + rS\frac{\partial \Pi}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \Pi}{\partial S^2} = r\Pi$$

Substituting for the definition of delta, gamma, and theta yields

$$\Theta + rs\Delta + \tfrac{1}{2}\sigma^2 S^2 \Gamma = r\Pi$$

From the differential equation, for a given delta, theta, and gamma must have an offsetting relationship. Thus, it would be impossible to eliminate delta, gamma, and theta risk completely unless the value of the portfolio is zero.

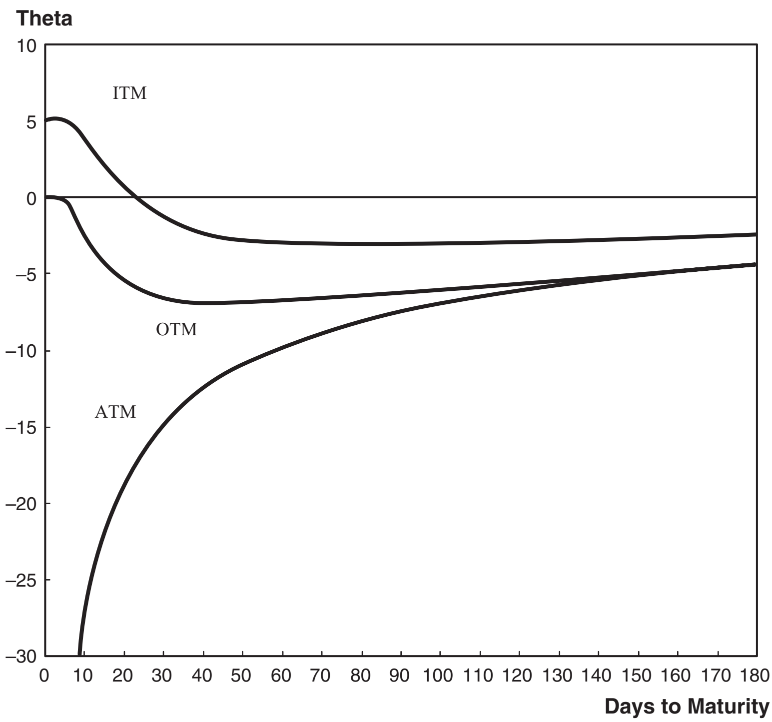


Exhibit 32.14 Put Thetas as a Function of Time to Maturity

VEGA

Vega is the rate of change of the option price relative to changes in the volatility of the stock. Even though vega is not a Greek letter, it is still considered one of the Greeks. Mathematically, the vega for a call option is

$$\begin{aligned} Vega_c &= \frac{\partial c}{\partial \sigma} \\ &= Se^{-qT}\sqrt{T}N'(d_1) \end{aligned} \tag{32.6}$$

The vega for a call option is identical to the vega for the corresponding put. Differentiating the put-call parity with respect to volatility yields

$$c - p = Se^{-qT} - Ke^{-rT}$$
$$\frac{\partial(c-p)}{\partial\sigma} = \frac{\partial(Se^{-qT} - Ke^{-rt})}{\partial\sigma}$$

Applying the definition of vega and noting that neither the stock price or the strike price is a function of volatility yields

$$\frac{\partial c}{\partial \sigma} - \frac{\partial p}{\partial \sigma} = 0$$

Thus,

$$\frac{\partial c}{\partial \sigma} = \frac{\partial p}{\partial \sigma}$$
$$Vega_c = Vega_p$$

Hence, $Vega_c = Vega_p = Vega$. Additionally, as with both delta and gamma, the vega for a portfolio is the sum of the vega exposure of each asset in the portfolio.

Example 5

For the parameters given in Exhibit 32.2, the vega of the call is 19.6103. What is expected price for the call option if the volatility increases to 26 percent?

The stock volatility rising from 25 to 26 percent is a 1 percent increase. Given a vega of 19.61, a 1 percent increase in the stock volatility should increase the price of the call by $19.6103 \times 0.01 = 0.1961$, or about 20 cents. The resulting expected price for the call is the original price of \$5.58 plus the \$0.20 for a total of \$5.78. The actual new price of the call option is \$5.77 for an error of less than 1 cent. Because vega is the same for puts and calls, the price of the corresponding put should also increase \$0.20.

The pattern for vega as a function of the stock price is similar to the pattern for gamma. Vega is the largest for near- or at-the-money options and decreases the farther one moves away from this level. Exhibit 32.15 graphs the relationship between the stock price and vega.

Vega is always positive. This can be seen mathematically as each term in Equation 32.6 must be positive. This results because at any point on a call functions

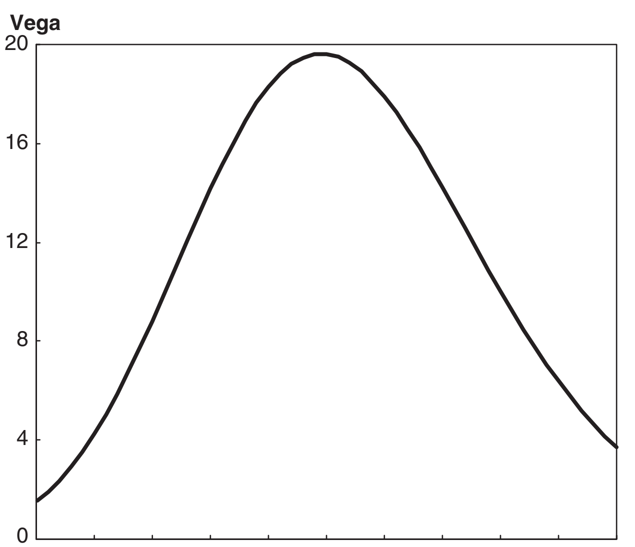


Exhibit 32.15 Vega as a Function of Stock Price

(like the one graphed in Exhibit 32.6), the value of the call increases at an increasing rate and decreases at a decreasing rate. Thus, increasing the volatility increases the upside potential more than the downside risk. The same argument can be made for puts.

Vega is the greatest for near the money options. This is because the asymmetry between the increasing upside potential and the decreasing downside risk is the greatest for these options. As puts and calls move away from this level, the payout function flattens out. Thus, the asymmetry between the increasing upside potential and the decreasing downside risk diminishes.

RHO AND OTHER OPTION SENSITIVITIES

Option prices are also sensitive the changes in the risk-free rate and the dividend yield. Rho is the rate of change in the option price relative to changes in the risk-free rate. Mathematically, rho for a call option is

$$\begin{aligned} Rho_c &= \frac{\partial c}{\partial r} \\ &= TKe^{-rT}N(d_2) \end{aligned} \tag{32.7}$$

For a put option, rho is

$$\begin{aligned} Rho_p &= \frac{\partial p}{\partial r} \\ &= -TKe^{-rT}N(-d_2) \end{aligned} \tag{32.8}$$

Because Greek capital letter for rho is ρ, which is the same symbol used to denote an American put option, the convention is to write out the word *Rho*.

Example 6

For the parameters given in Exhibit 32.2, the rho of the call is 12.65 and the rho for the put is -11.91. What are the expected prices for the call and put options if the risk-free rate increases to 6.25 percent?

The risk-free rate increasing from 6.00 to 6.25 percent is a 25 basis point increase. Given a rho for the call of 12.65, a 25 basis point increase in the risk-free rate should increase the price of the call by $12.65 \times 0.0025 = 0.03$ or about 3 cents from \$5.58 to \$5.61. In the case of the put, the 25 basis point increase in the risk-free rate should cause the price to drop by $\$0.03 = -11.91 \times 0.0025$. This corresponds to a new expected put price of $\$4.34 - \$0.03 = \$4.31$. The actual new prices for the call and put options that correspond to a risk-free rate of 6.25 percent are \$5.61 and \$4.31.

For both the put and the call, the effect of rho increases as the stock price increases. Exhibit 32.16 graphs the relationship between rho and the stock price.

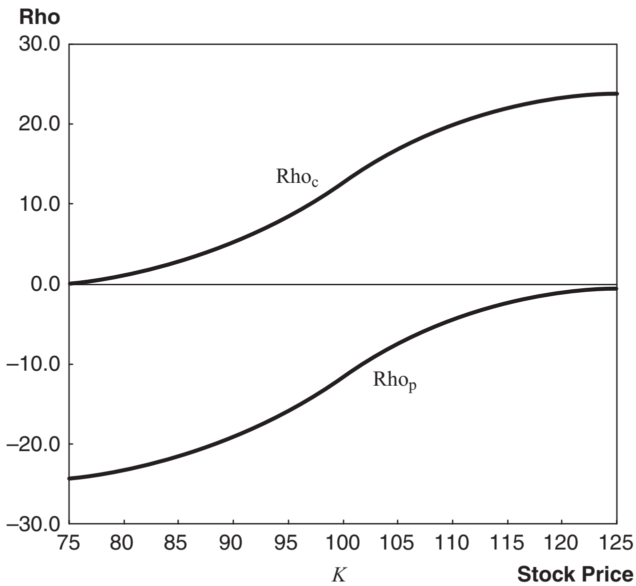


Exhibit 32.16 Rho as a Function of Stock Price

Psi is the rate of change in the option price relative to changes in the dividend yield rate. Mathematically, the psi of a call option is

$$\begin{aligned}\Psi_c &= \frac{\partial c}{\partial q} \\ &= -TSe^{-qT}N(d_1)\end{aligned} \tag{32.9}$$

For a put option, psi is

$$\begin{aligned}\Psi_p &= \frac{\partial p}{\partial q} \\ &= TSe^{-qT}N(-d_1)\end{aligned} \tag{32.10}$$

For both the put and the call, the effect of psi decreases as the stock price increases. Exhibit 32.17 graphs the relationship between psi and the stock price.

Example 7

For the parameters given in Exhibit 32.2, the psi of the call is −14.04 and the psi for the put is 10.83. What are the expected prices for the call and put options if the dividend yield increases to 1.25 percent?

The dividend yield increasing from 1.00 to 1.25 percent is a 25 basis point increase. Given a psi for the call of −14.04, a 25 basis point increase in the dividend yield should change the price of the call by $-14.04 \times 0.0025 = 0.0351$, or about 4 cents from \$5.58 to \$5.54. In the case of the put, the 25 basis point increase in the dividend yield should cause the price to increase by \$0.0271 $= 10.83 \times 0.0025$.

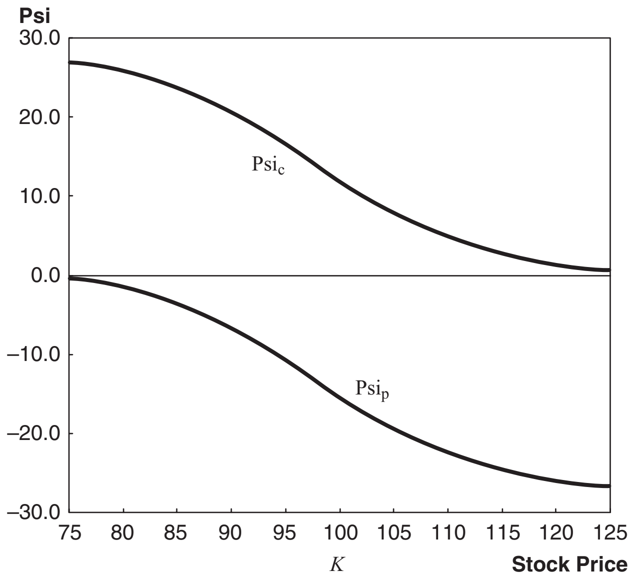


Exhibit 32.17 Psi as a Function of Stock Price

This corresponds to a new expected put price of \$4.34 + \$0.03 = \$4.34. The actual new prices for the call and put options that correspond to a dividend yield of 1.25 percent are \$5.54 and \$4.37.

In the case of a foreign currency option, the dividend yield q in the Merton model can be replaced by the foreign risk-free rate r_{for}. The option pricing model becomes

$$\begin{aligned} c &= Se^{-r_{for}T}N(d_1) - Ke^{-rT}N(d_2) \\ p &= Ke^{-rT}N(-d_2) - Se^{-r_{for}T}N(-d_1) \\ d_1 &= \frac{\ln(S/K) + (r - r_{for} + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \\ d_2 &= d_1 - \sigma\sqrt{T} \end{aligned} \tag{32.11}$$

where r_{for} = foreign risk-free rate

For foreign currency options, pricing sensitivity to the foreign risk-free rate is identical to the sensitivity to the dividend yield in the Merton model. Thus, for a call

$$\begin{aligned} ForRho_c &= \frac{\partial c}{\partial r_{for}} \\ &= -TSe^{-r_{for}T}N(d_1) \end{aligned} \tag{32.12}$$

and for a put

$$\begin{aligned} ForRho_p &= \frac{\partial p}{\partial r_{for}} \\ &= TSe^{-r_{for}T} N(-d_1) \end{aligned} \tag{32.13}$$

where $d_1 =$ is from Equation 32.11

Because the foreign interest rate is a risk-free rate, it is customary to refer to this sensitivity as rho foreign.

Theoretically, options prices are also sensitive to changes in the strike price. Mathematically, the sensitivity of a call price to changes in the strike price is

$$\frac{\partial c}{\partial K} = -e^{-rT} N(d_2) \tag{32.14}$$

For a put, the relationship is

$$\frac{\partial p}{\partial K} = e^{-rT} N(-d_2) \tag{32.15}$$

However, for most options, the strike price is fixed over the life of the option. This is probably why this particular sensitivity does not have a standard Greek nomenclature. Additionally, this sensitivity is more useful in applications where one is comparing similar options across a range of strike prices rather than examining the sensitivity of a single option to strike price risk.

HEDGING DELTA, GAMMA, AND VEGA

As noted in Examples 6 and 7, reasonable changes in the risk-free rate and the dividend yield do not have a large effect on put and call prices. Additionally, the changes in the risk-free rate or the dividend yield are usually small. Also, for most options the strike price does not change over the life of the option. This leaves delta, gamma, vega, and theta risk. Since the effects of time decay in options are fairly predictable, we will look at hedging delta, gamma, and vega risk. Please see Exhibit 32.18.

Assume that you have a portfolio of stock options (all on the same stock) with in exposures given in Exhibit 32.18. The first step to managing the delta, gamma, and vega risk of the portfolio is to measure the exposure to each risk. As noted

Exhibit 32.18 Options to Hedge

Type	**Position**	**Delta**	**Gamma**	**Vega**
Call	100	0.60	0.15	3.40
Call	−300	0.40	0.33	1.90
Put	200	−0.30	0.18	2.05
Put	−400	−0.75	0.15	4.70

earlier, the exposure for delta is sum of the delta exposure of each asset in the portfolio. Gamma and vega exposure are calculated similarly. Thus,

$$\begin{aligned}
\Delta_P &= 100 \times 0.60 + -300 \times 0.40 + 200 \times -0.30 + -400 \times -0.75 \\
&= 180 \\
\Gamma_P &= 100 \times 0.15 + -300 \times 0.33 + 200 \times 0.18 + -400 \times 0.15 \\
&= -108 \\
Vega_P &= 100 \times 3.40 + -300 \times 1.90 + 200 \times 2.05 + -400 \times 4.70 \\
&= -1700
\end{aligned}$$

To hedge delta and gamma risk, assets will have to be added to the portfolio that have a delta of −180 and a gamma of 108. Exhibit 32.19 gives the exposure for two additional options that are available for hedging. Also, as discussed in prior sections, adding a share of stock will increase the delta of the portfolio by 1 but will leave the gamma and vega unaltered.

Because the first option has a gamma of 0.04, we need to add $108/0.04 = 2{,}700$ of these options to make the portfolio gamma-neutral. If we add 2,700 of the first option to the portfolio, this will increase the delta of the portfolio by $2{,}700 \times 0.5 = 1{,}350$. The delta for the portfolio would then be the old delta of 180 plus the delta of 1,350 from the 2,700 new options for a resulting delta of 1,530. This position can be made delta-neutral by shorting 1,530 shares of the stock. Because the gamma for a share of stock is zero, the gamma of the portfolio will be unaltered. Thus, the portfolio represented in Exhibit 32.2 can be made delta- and gamma-neutral by adding 2,700 of the first option and shorting 1,530 shares.

To make the portfolio delta- and vega-neutral, assets will have to be added that have a delta of −180 and a vega of 1,700. The first option has a vega of 4.25. Thus, adding 400 of the first option will increase the vega of the portfolio by $400 \times 4.25 = 1{,}700$. This will make the portfolio vega-neutral. However, adding 400 of the first option will increase the delta of the portfolio by $400 \times 0.5 = 200$. The new delta of the portfolio would be the delta from the original portfolio of 180 plus 200 due to the addition of the new options for a new delta of 380. Shorting 380 shares of stock will then make the portfolio delta-neutral. Because the vega of a share of stock is zero, adding 380 shares of stock to the portfolio leaves the vega of the portfolio unaltered. Thus, the portfolio represented in Exhibit 32.2 can be made delta- and vega-neutral by adding 400 of the first option and shorting 380 shares.

To make the portfolio delta-, gamma-, and vega-neutral, we need to use both of the additional options listed in Exhibit 32.19. Denoting the number of the first option as Opt_1 and the number of the second option as Opt_2, we can set up a system of two equations in two unknowns. For the portfolio to be gamma- and

Exhibit 32.19 Additional Options Available for Hedging

	Delta	Gamma	Vega
Opt_1	0.50	0.04	4.25
Opt_2	0.20	0.02	5.00
Stock	1.00	0.00	0.00

vega-neutral, the two options will have to add 108 to the gamma position and add 1,700 to the vega position. Setting up one equation for delta and one equation for gamma gives

$$\Gamma : 108 = Opt_1 \times 0.04 + Opt_2 \times 0.02$$
$$Vega : 1{,}700 = Opt_1 \times 4.75 + Opt_2 \times 5.00$$

Solving this system of two equations in two unknowns yields

$$Opt_1 = 4{,}400$$
$$Opt_2 = -3{,}400$$

With the addition of 4,400 of Opt_1 and $-3{,}400$ of Opt_2, the delta of the portfolio will change to

$$180 + Opt_1 \times 0.5 + Opt_2 \times 0.2 =$$
$$180 + 4{,}400 \times 0.5 + 1{,}700 \times 0.2 = 1{,}700$$

To make this new position delta-neutral, one would need to short 1,700 shares of stock. Again, because the gamma and vega for a share of stock are zero, shorting 1,700 shares of stock maintains the gamma and vega neutrality of the portfolio. Thus, adding 4,400 of the first option, shorting 3,400 of the second option, and shorting 1,700 shares of stock would make the portfolio in Exhibit 32.18 delta-, gamma-, and vega-neutral.

CONCLUSION

This chapter analyzes how to measure and hedge pricing risks for the key inputs to the Merton option pricing model. These inputs are the stock price, stock price volatility, the time to maturity, the risk-free rate, the dividend yield, and the strike price. Risks associated with these inputs are delta, gamma, vega, theta, rho, psi, and a foreign rho. Delta is the option sensitivity to changes in the stock price while gamma measures how delta changes when the stock price changes. The time decay of an option is measured by theta, and vega measures the option price sensitivity to changes in the stock price volatility. Rho and psi are the sensitivities to the risk-free rate and dividend yield, respectively. In the case of foreign currency options, the price sensitivity to the foreign risk-free rate is known as the foreign rho.

Additionally, this chapter analyzes strategies for hedging against the risks that are the most likely to have the greatest impact on the performance of a portfolio of options: delta, gamma, and vega. Hedging a portfolio of options against delta and gamma risk requires taking a position in an additional traded option and the stock. Delta and vega can be hedged simultaneously in a similar manner. However, hedging delta, gamma, and vega requires taking a position in two additional traded options and the stock.

REFERENCES

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, no. 3 (May/June): 637–654.

Cox, J. C., S. A. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach," *Journal of Financial Economics* 7, no. 3 (October): 229–263.

Merton, R. C. 1973. "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4, no. 1 (Spring): 141–183.

ABOUT THE AUTHOR

R. Brian Balyeat is an assistant professor of finance at Xavier University. He earned a BA in mathematics and economics in 1989 from Rhodes College, an MBA from Duke University in 1994, and a PhD in finance from Duke University in 1998. Prior to entering academia, he was a financial analyst for First Tennessee Bank for three years, where he earned his Chartered Financial Analyst designation.

PART VI

Using Financial Derivatives

The enormous size and activity of the world's derivatives markets attest to the fact that millions of individuals, firms, and governments find them very useful. As earlier chapters have emphasized derivatives markets have legitimate and valid social functions. This section discusses some of the ways that market participants use these powerful instruments.

Stewart Mayhew, in Chapter 33, "Option Strategies," reviews the mechanics of trading strategies involving portfolios of options and their underlying securities to provide an overview of the combinations, straddles, and spreads that make up the standard repertoire of option strategies. Mayhew notes that one key use of options is as a vehicle for traders to take positions reflecting their views about the future volatility of underlying instruments, and he shows how options can be used to buy and sell volatility. As Mayhew discusses, this ability to use financial derivatives to shape risk and return exposure to match the desires of a particular investor is one of the great attractions and uses of derivatives markets.

In Chapter 34, "The Use of Derivatives in Financial Engineering: Hedge Fund Applications," John F. Marshall and Cara M. Marshall begin by defining financial engineering and presenting an overview of how firms and portfolio managers use the techniques of financial engineering. Against this general background, the Marshalls turn to specific examples of financial engineering in hedge funds. The chapter mainly focuses on illustrating financial engineering by showing how these engineering techniques can be used in convertible bond arbitrage and capital structure arbitrage.

Hedge funds have become major users of financial derivatives, and some observers see the tremendous growth in credit derivatives (that is, the tremendous growth leading up to the financial crisis) as being driven largely by hedge funds. Tom Nohel explores the role of hedge funds in the derivatives markets and the uses that hedge funds make of financial derivatives in Chapter 35, "Hedge Funds and Financial Derivatives." For example, Nohel shows how hedge funds use derivatives trading strategies to implement their key investment decisions in ways that economize on transaction costs.

As the financial crisis has developed, interest rates have been driven ever lower. To some, this may suggest that interest rate risk is of less importance than in previous times. However, it may well be that interest rates will quickly surge as the crisis subsides, particularly given the massive injections of liquidity provide by central banks and governments around the world. As Steve Byers notes in Chapter 36, "Using Derivatives to Manage Interest Rate Risk," financial derivatives are particularly potent tools for managing interest rate risk. Byers classifies interest rate risk management techniques by the type of derivatives instruments, focusing first on forward-based instruments (futures and forward contracts) and then turning to interest rate options. In the course of his discussion he explains interest rate caps, floors, and collars.

Betty J. Simkins and Kris Kemper take up the question of real options in Chapter 37, "Real Options and Applications in Corporate Finance." They define real options as ". . . option-like opportunities such as business decisions and flexibilities, where the underlying assets are real assets . . ." For example, the opportunity to expand into a new market is like a more typical option in that it has a value, the value fluctuates with changes in economic conditions, and it ultimately comes to expiration and must be exercised or it expires worthless—that is, it ceases to be an opportunity. More and more, finance scholars and corporate managers are realizing that the real options approach to analyzing problems provides a very powerful way of thinking about opportunities and decisions, and Simkins and Kemper make the case for the importance of these real options.

CHAPTER 33

Option Strategies

STEWART MAYHEW
Deputy Chief Economist, U.S. Securities and Exchange Commission

It has long been known that, in theory, a portfolio of ordinary European call or put options can be used to complete the market. (See Ross 1976.) With frictionless markets and a full set of contracts, virtually any desired payoff function can be achieved with the appropriate combination of purchased and written options. In practice, the extent to which investors can use portfolios of options to produce customized payoff functions is limited by the sparsity of strike prices and maturities available for contracting and by the transactions costs associated with implementing the strategies.

This chapter reviews the mechanics of simple trading strategies involving portfolios of options and their underlying securities. Its main purpose is to provide an overview of the combinations, straddles, and spreads that make up the standard repertoire of option strategies, not to evaluate the merits of any particular strategy. Nevertheless, it may be worthwhile at the outset to enumerate a few basic motivations underlying these option strategies, to emphasize that among the many strategies outlined, different ones are likely to be used by different types of traders.

One key use of options is as a vehicle for traders to take positions reflecting their views about future volatility. The option premium reflects the market's assessment of volatility in the underlying asset over the remaining life of the option. Those who believe that future volatility will be higher than what is reflected in option prices—or those who believe the market's forecast of volatility is likely to increase—may wish to "buy" volatility by buying calls and/or puts. Those who trade volatility in this manner generally will find that near-the-money calls and puts are the best vehicles, as options that are very far in- or out-of-the money are not very sensitive to volatility. The investor might overweight calls or puts in order to combine the volatility bet with a directional bet, or take a pure volatility position by hedging out the delta (the risk exposure to the underlying stock).[1]

A second motivation for trading options is as a substitute vehicle for taking on exposure to the underlying stock. Buying a call option allows the investor to get

---

The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The views expressed herein are those of the author and do not necessarily reflect the views of the Commission or of the author's colleagues upon the staff of the Commission.

levered upside exposure, much like buying the underlying security on margin, but with limited downside risk. Likewise, buying put options is a substitute vehicle for short selling the underlying stock that allows the investor levered negative exposure while limiting losses if the underlying should increase. A long option position is equivalent to a dynamic trading strategy in the underlying stock for which the risk exposure automatically decreases as the position loses money.

In addition to this nonlinear risk profile, there are various other reasons why investors may prefer options as a vehicle for trading exposure to the underlying stock. Investors with access to good information may seek to spread their trades across multiple trading venues in an effort to disguise their intentions.[2] Investors who face capital constraints may use options to achieve a degree of leverage that would not be achievable by trading in the underlying market. Investors may look to option markets in cases where the underlying market has abnormally high transactions costs (e.g., due to inefficient clearing and settlement mechanisms). Investors may seek to trade options if there are regulatory restrictions in the underlying market that do not apply in the option market, such as an uptick rule or a requirement to locate shares before short selling. Some investors have attempted to use options to circumvent other restrictions on short selling.[3] Investors may wish to trade options rather than underlying assets if the provisions of the tax code create incentives to do so.

Another motivation for trading options is to hedge or alter the risk profile of existing positions in the underlying security. An investor who owns a stock but wants to neutralize risk exposure for a limited period of time can hedge the stock position with options and avoid the tax implications and logistical problems associated with selling and repurchasing a large block of stock. Investors wishing to insure specific stock positions against large losses can buy protective puts, and those wishing to sell off the upside potential of specific positions for cash today can write covered calls.

Finally, arbitrageurs may wish to trade options to exploit opportunities when options are temporarily mispriced relative to each other or relative to the underlying stock and default-free bonds. It is well known that option prices must satisfy various properties in order to prevent arbitrage, such as put call parity and the other restrictions discussed by Merton (1973). When these are violated, option traders can profit by trading the appropriate portfolio of options and stock. More generally, traders with proprietary pricing models may identify opportunities that, though not perfect arbitrage, are "good deals."[4] These traders, perceiving an option to be mispriced, would trade the option and dynamically hedge the position.

In sum, there are numerous reasons why traders may wish to trade options. Of course, the attractiveness of a particular strategy at a particular time will depend on the goals of the trader, the prices of the options and the underlying stock, the interest rate, and the transaction costs associated with implementing the strategy. The reader of this chapter will become familiar with an array of simple option strategies, will understand how the component pieces of these strategies fit together, and should have the conceptual toolkit necessary to analyze other, more complicated strategies.

I begin by reviewing the concept of payoff tables and profit/loss diagrams, using them to illustrate the four basic option positions. I then move on to discuss protective puts and covered calls. Then I show how options can be used to create synthetic forward positions, how stock and options can be used to create synthetic borrowing and lending through "conversions" and "reversals," how calls can used to create synthetic puts, and how puts can be used to create synthetic calls. Next I describe simple bull and bear spreads and positions that are economically similar to them (collars and range forwards). Then I describe a position sometimes called a cylinder, which is basically a synthetic forward but with multiple strike prices. Straddles, strangles, and other positions that generally would be used to take volatility positions are discussed next followed by ratio spreads, box spreads, and butterfly spreads and other related positions. Although the primary focus of this chapter is on strategies involving options that are in the same class and that expire simultaneously, next I touch on strategies involving options at multiple maturities and very briefly discuss strategies involving multiple underlying assets, including correlation-sensitive strategies.

Before proceeding, one final warning. In this chapter, I glibly assign names to strategies, as if there existed a universally accepted nomenclature. In fact, different trading communities often have different names for the same strategies, especially when going beyond the standard textbook positions. So, when I refer to strategies with colorful names such as the Christmas tree, the seagull, or the albatross, be aware that the terminology may not be universal.

BUILDING BLOCKS

This chapter describes a variety of strategies involving combinations of stock, cash, and European-style options. In general, the same strategies also may be implemented with American-style options, with the caveat that the strategies then will be susceptible to being undermined by early exercise, if they involve written option positions.[5] I focus attention on the performance of these strategies over a horizon corresponding to the expiration of the options. I do not address issues related to the short-term risk characteristics of these strategies, such as the portfolio Greeks, as this topic is addressed elsewhere.

The key to understanding and analyzing portfolios of options is understanding how to aggregate payoff functions across multiple positions in the same underlying asset. The strategies to be discussed are built from combinations of eight basic components: long and short positions in the stock, purchased and written positions in calls, purchased and written positions in puts, and borrowing and lending cash. I describe the strategies using payoff tables and profit/loss diagrams.

The payoff table defines different states of the world based on the level of the underlying stock price when the options expire, relative to the strike prices of all the options used in the strategy. To illustrate with the simplest possible strategy, consider a purchased position in a European-style call option. If S represents the stock price at expiration and K represents the strike price, payoff table would be constructed as follows:

Payoff Table for a Purchased Call

	$S < K$	$K < S$
Purchased K call	0	$S - K$

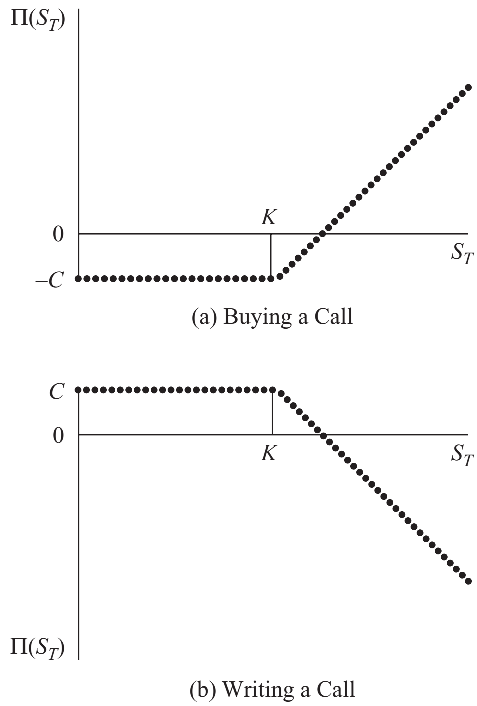


Exhibit 33.1 Profit/Loss Diagrams for Call Option Positions

For a written call option, the payoff table would be:

Payoff Table for a Written Call

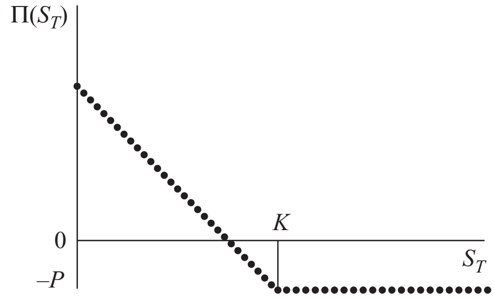
	$S < K$	$K < S$
Written K call	0	$K - S$

The profit/loss diagram graphs the total payoff of the strategy (as described in the payoff table), adjusted for the cashflow associated with entering the strategy, as a function of the terminal stock price. In other words, it is simply the payoff function, shifted down (up) to reflect the cash outflow (inflow) at the initiation of the strategy. The profit/loss diagram for purchased and written calls is illustrated in Exhibit 33.1 where C represents the price of the call option.[6]

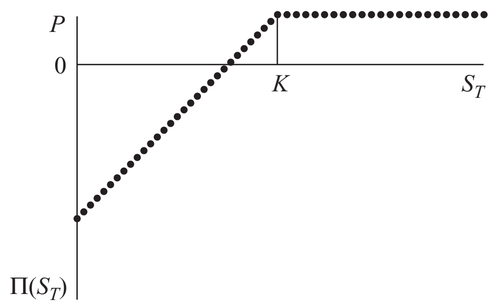
Likewise, the payoff tables for the put option are shown in the next tables and in Exhibit 33.2.

Payoff Table for a Purchased Put

	$S < K$	$K < S$
Purchased K put	$K - S$	0



(a) Buying a Put



(b) Writing a Put

Exhibit 33.2 Profit/Loss Diagrams for Put Option Positions

Payoff Table for a Written Put

	$S < K$	$K < S$
Written K put	$-(K - S)$	0

I have now described four basic positions involving options: buying a call, writing a call, buying a put, and writing a put. In addition, the investor may borrow or lend cash and buy or short the underlying stock, bringing us to eight basic positions. Keep in mind that on any particular stock, call and put options with various strike prices and maturities are available, and an investor may simultaneously take multiple positions in calls, puts, the underlying stock, and cash. Thus, a wide variety of different combinations are possible.

COVERED CALLS AND PROTECTIVE PUTS

This section discusses simple strategies involving the stock and one option. A covered call strategy involves a long position in the underlying combined with a written call option. Simply stated, an investor who writes a covered call is selling off the underlying asset's upside potential in exchange for cash today.

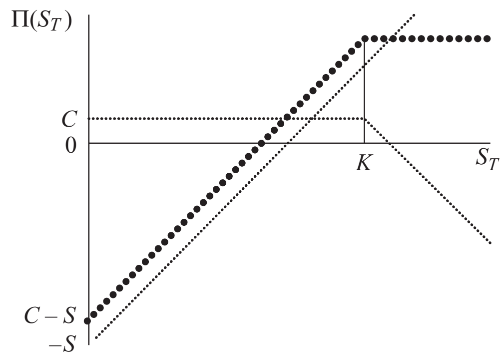
The payoffs for a covered call position are summarized next.

Payoff Table for a Covered Call

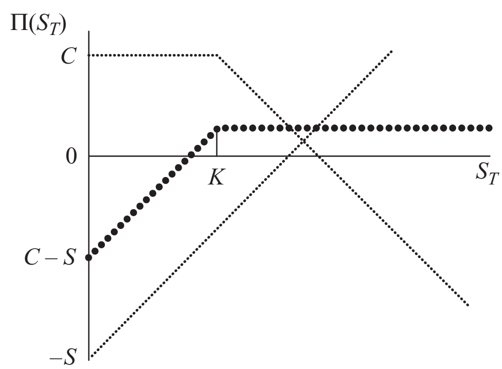
	$S < K$	$K < S$
Long stock	S	S
Written K call	0	$-(S - K)$
Total	S	K

Exhibit 33.3 shows the profit/loss function for the covered call position (heavy dotted line) and the contribution of each component (light dotted lines). Exhibit 33.3(a) depicts the profit/loss diagram for writing an out-of-the-money covered call, and Exhibit 33.3(b) shows the same for writing an in-the-money covered call.

Buying a put option on a stock that is held long is called buying a protective put. This is like buying an insurance policy that covers any losses incurred as a result of the stock falling below the strike price.



(a) Writing a Covered Call—High Strike Price



(b) Writing a Covered Call—Low Strike Price

Exhibit 33.3 Profit/Loss Diagrams for Covered Call Option Positions

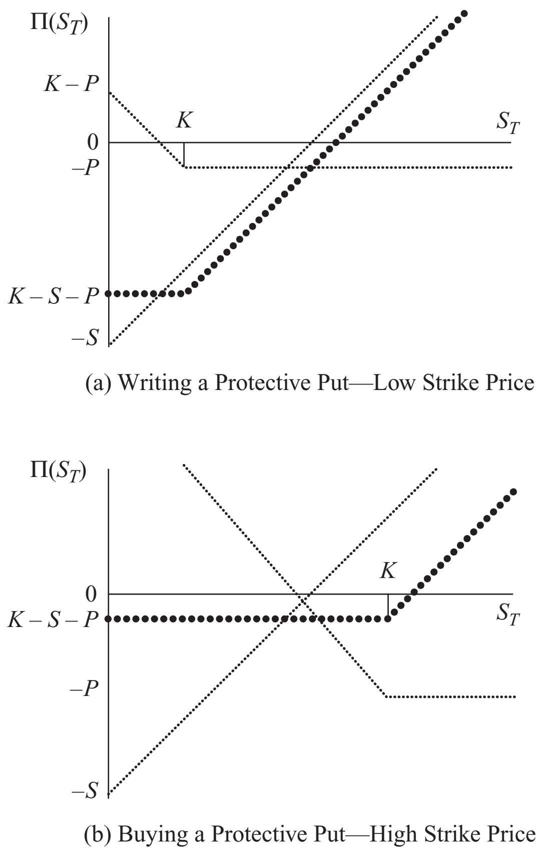


Exhibit 33.4 Profit/Loss Diagrams for Protective Put Option Positions

Payoff Table for a Protective Put

	$S < K$	$K < S$
Long stock	S	S
Purchased K put	$K - S$	0
Total	K	S

Exhibit 33.4 shows how the profit/loss function for a protective put is decomposed into the stock and put positions and how the payoff diagram differs as a function of the strike price. At a low strike price, the put option is cheap but does not offer as much protection.

SYNTHETIC POSITIONS

The relation among call options, put options, the underlying asset, and risk-free bonds is summarized in the put-call parity formula, which implies that any one of these assets can be created synthetically from the others. This section shows how options can be combined to create synthetic forwards, how options and the underlying asset can be combined to create synthetic borrowing and lending, how the underlying asset and calls can be combined to create a synthetic put, and how the underlying asset and puts can be combined to create synthetic calls.

A forward contract locks in a price today for a transaction in the future, where the long party agrees to buy and the short side agrees to sell the underlying asset for a predetermined price, called the delivery price. The forward price is the delivery price such that both parties are willing to agree to enter the contract with no initial payment on either side. If the delivery price is set higher (lower) than the forward price, the party on the long (short) side would not be willing to enter the contract without additional cash compensation.

If K represents the delivery price and S is the underlying asset price when the forward expires, the payoff for the long side of a forward is simply $S - K$, and the payoff for the short side is $K - S$. An equivalent position can also be created with options. A long synthetic forward is created by buying a call and writing a put with the same strike price, K.

Payoff Table for a Long Synthetic Forward

	$S < K$	$K < S$
Purchased K call	0	$S - K$
Written K put	$S - K$	0
Total	$S - K$	$S - K$

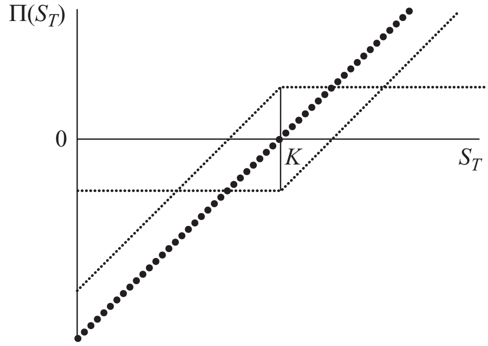
A short forward position can by synthesized by writing a call and buying a put, with strike prices equal to the desired delivery price.

Payoff Table for a Short Synthetic Forward

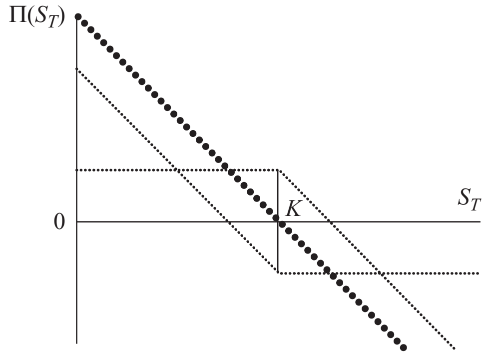
	$S < K$	$K < S$
Written K call	0	$-(S - K)$
Purchased K put	$K - S$	0
Total	$K - S$	$K - S$

If the strike price is chosen to be equal to the forward price, then the put price should equal the call price, and no initial investment should be required. If a higher strike price is chosen, the put price will exceed the call price, and the long (short) synthetic forward will generate cash inflow (outflow). If the strike is below the forward price, the opposite is true. Exhibit 33.5 depicts profit/loss diagrams for long and short synthetic forward positions constructed at the forward price.

Consider a repurchase agreement where the initial purchase and subsequent sale are negotiated simultaneously. Or equivalently, consider the strategy of buying an asset today and at the same time locking in a future selling price with a forward contract. The return on this strategy does not depend on what happens to the



(a) Long Synthetic Forward Position



(b) Short Synthetic Forward Position

Exhibit 33.5 Profit/Loss Diagrams for Synthetic Forward Positions

underlying asset price. In fact, this transaction is equivalent to lending money. Likewise, short selling an asset and simultaneously locking in a future buying price is equivalent to borrowing money. The synthetic borrowing rate will be determined by the spot price and the forward price.

Just as a forward or futures contract may be used to create a synthetic lending or borrowing position, a synthetic forward made of options may also be used for this purpose. The synthetic lending position, known as a conversion strategy, would include buying the stock, writing a call, and buying a put option with the same strike price K.

Payoff table for a Conversion (Synthetic Lending)

	$S < K$	$K < S$
Long stock	S	S
Written K call	0	$K - S$
Purchased K put	$K - S$	0
Total	K	K

Synthetic borrowing, or a "reversal" strategy, is exactly the opposite position: short the underlying stock, a written put option, and a purchased call.

Payoff table for Reversal (Synthetic Borrowing)

	$S < K$	$K < S$
Short stock	$-S$	$-S$
Purchased K call	0	$S - K$
Written K put	$S - K$	0
Total	$-K$	$-K$

A put option can be synthesized by short selling the stock, buying a call option with strike price K, and investing the present value of K at the risk-free rate.

Payoff Table for a Synthetic Put

	$S < K$	$K < S$
Short stock	$-S$	$-S$
Purchased K call	0	$S - K$
Lending	K	K
Total	$K - S$	0

Likewise, writing a covered call at strike price K and borrowing the present value of K is equivalent to writing a put. (Compare the profit/loss diagrams and note that the payoff of a covered call differs from that of a written put only by a constant.)

A purchased call is equivalent to a portfolio including a long position in the stock, a put option, and borrowed cash:

Payoff Table for a Synthetic Call

	$S < K$	$K < S$
Long stock	S	S
Purchased K put	$K - S$	0
Borrowing	$-K$	$-K$
Total	0	$S - K$

BULL AND BEAR SPREADS

This section describes different versions of bull spreads and bear spreads. These positions enable an investor to take directional bets on the underlying stock but limit the risk exposure to a defined region of stock prices. In other words, the position has full exposure to small stock price changes but a cap on both total losses and on total gains.

A bull spread involves simultaneously buying a call option with a lower strike price and writing a call with a higher strike price, where the two options have the same expiration date. Because the call with the lower strike will have the higher price, the bull spread requires an initial cash investment.

When the options expire, there are three possible scenarios: The underlying stock price may end up below the lower strike price, in between the two, or above the higher strike price. In the first case, both options finish out-of-the-money, and the trader loses the entire initial investment. In the second case, the purchased option will finish in-the-money, but the written option is out-of-the-money. In this case, the trader recovers some or all of the initial investment and may even make some money. In the final case, both options finish in-the-money. The trader buys the underlying stock at a low price by exercising the purchased option, then sells it at a higher price because the written option is exercised. The payoff at expiration is equal to the difference between the two strike prices.

Payoff Table for a Bull Spread

	$S < K_L$	$K_L < S < K_H$	$K_H < S$
Purchased K_L call	0	$S - K_L$	$S - K_L$
Written K_H call	0	0	$-(S - K_H)$
Total	0	$S - K_L$	$K_H - K_L$

A bear spread involves buying a put option with a high strike price and writing a put option with a low strike price. A bear spread also requires an initial investment.

Payoff Table for a Bear Spread

	$S < K_L$	$K_L < S < K_H$	$K_H < S$
Written K_L put	$-(K_L - S)$	0	0
Purchased K_H put	$K_H - S$	$K_H - S$	0
Total	$K_H - K_L$	$K_H - S$	0

Exhibit 33.6 illustrates the decomposition of the profit/loss function for the bull spread and the bear spread.

Now consider the position taken by the counterparty to the bear spread, who has purchased a put with a low strike price and written a put with a high strike price. This position has the same profile as a bull spread but generates initial income and a subsequent liability. It is known as a credit bull spread. Similarly, the position opposite the bull spread is known as a credit bear spread.

Another position that has the same payoff profile as a bull spread is achieved by buying the underlying asset, buying a put option at a lower strike price, and

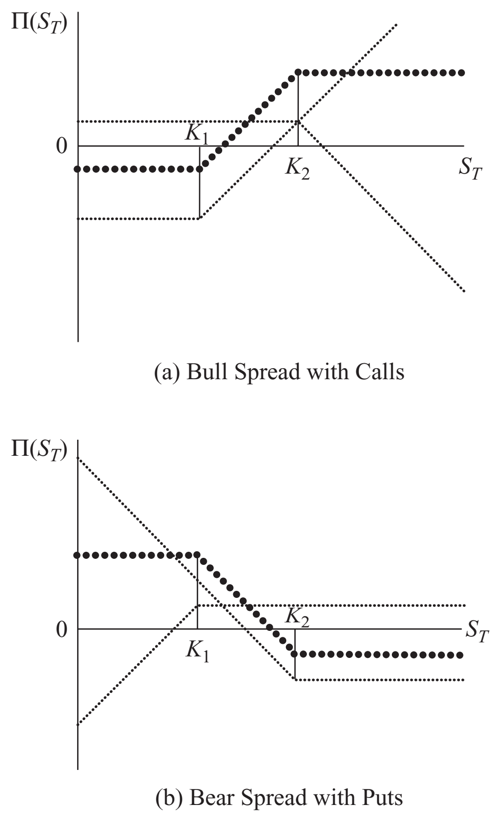


Exhibit 33.6 Profit/Loss Diagrams for Spread Positions

writing a call option at a higher strike price. This position, sometimes called a collar or a fence, requires more capital than the bull spread—in fact, it is equivalent to a bull spread plus cash. If one happens to already own the underlying asset, it is a relatively simple matter to convert a standard long position in the asset to a collar. Essentially, the investor is just buying a protective put and writing a covered call on the same asset. This is a quick way to reduce the risk of large price movements while still maintaining full exposure to local price changes.

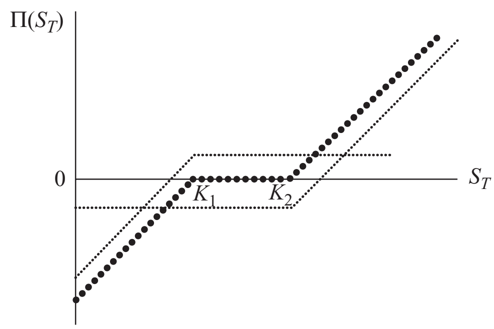
Payoff Table for a Collar

	$S < K_L$	$K_L < S < K_H$	$K_H < S$
Long stock	S	S	S
Purchased K_L put	$K_L - S$	0	0
Written K_H call	0	0	$-(S - K_H)$
Total	K_L	S	K_H

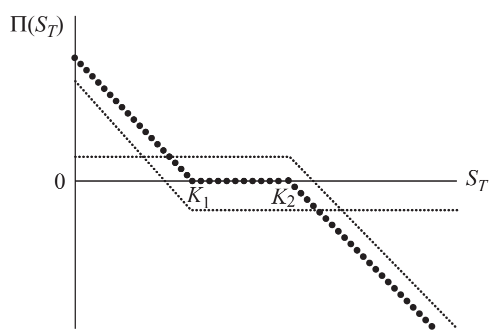
Another closely related position is achieved by using a forward instead of buying the asset. That is, one simultaneously enters the long side of a forward, buys a put option at the lower strike price, and writes a call option at the higher strike. One advantage of doing it this way is that by carefully choosing the two strike prices, this position can be constructed to have an initial cost of zero. Because the initial cost of entering a forward is zero, all that is necessary to accomplish this is to choose strike prices so that the two options have the same price. This position is called a range forward, and it is equivalent to a bull spread or a collar financed by borrowing.

CYLINDERS

Recall that a synthetic forward is constructed by simultaneously buying a call option and writing a put option at the same strike price. A cylinder is similar to a synthetic forward but with the put option written at a lower strike price. The profit/loss diagram of this position is illustrated in Exhibit 33.7. This position is similar to a forward but is less bullish—compared to the forward, the cylinder pays off less when the market goes up and more when the market goes down. In fact, a cylinder is equivalent to a forward plus a bear spread. In Exhibit 33.7, the strike prices were chosen so that the initial cost of buying the call is offset by the revenue from selling the put. In this case, the initial cost of the cylinder is zero. At lower strike prices, cylinders require an initial investment, while at higher strike prices, they produce income.



(a) Long Cylinder



(b) Short Cylinder

Exhibit 33.7 Profit/Loss Diagrams for Cylinders

STRADDLES, STRANGLES, STRIPS, AND STRAPS

This section describes a set of strategies that might be used by traders who wish to take positions with respect to volatility.

The simplest of these is the straddle, which involves simultaneously buying a call option and a put option at the same strike price and the same expiration. Typically, one would choose a strike price close to the money. Buying a straddle is a way to bet that there will be a large price change (or high volatility) in the stock without having to specify whether it will go up or down.

Payoff Table for a Straddle

	$S < K$	$K < S$
Purchased K call	0	$S - K$
Purchased K put	$K - S$	0
Total	$K - S$	$S - K$

The profit/loss diagram for buying a straddle is shown in Exhibit 33.8(a).

A strangle is another way to bet on high volatility. It is cheaper than a straddle but requires a larger move in the underlying stock before it pays off. To buy a

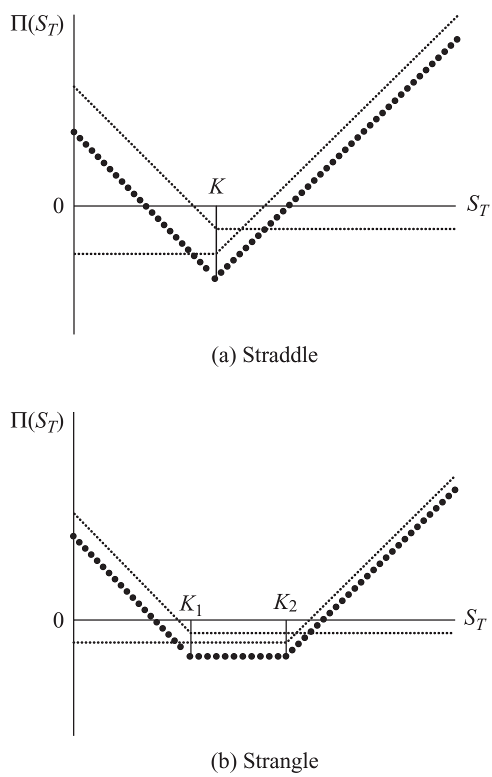


Exhibit 33.8 Profit/Loss Diagrams for Straddle and Strangle Positions

strangle, one would buy a call option with a higher strike price and a put option with a lower strike.

Payoff Table for a Strangle

	$S < K_L$	$K_L < S < K_H$	$K_H < S$
Purchased K_L put	$K_L - S$	0	0
Purchased K_H call	0	0	$S - K_H$
Total	$K_L - S$	0	$S - K_H$

The profit/loss diagram is shown in Exhibit 33.8(b).

Straddles that are purchased near the money may be considered pure volatility bets. In other words, these positions have positive payoffs if there is a large move in the underlying asset, but because the payoff function is symmetric, it makes no difference which direction the price moves. Suppose one believes that volatility will be high but also that the price is more likely to increase than decrease. This investor might consider buying a strap, consisting of one purchased put and two purchased calls at the same strike price. A bearish investor might prefer a strip, made up one call and two puts. Profit/loss diagrams for straps and strips are depicted in Exhibit 33.9. Of course, other ratios are possible, and one could modify the straddle in a similar fashion.

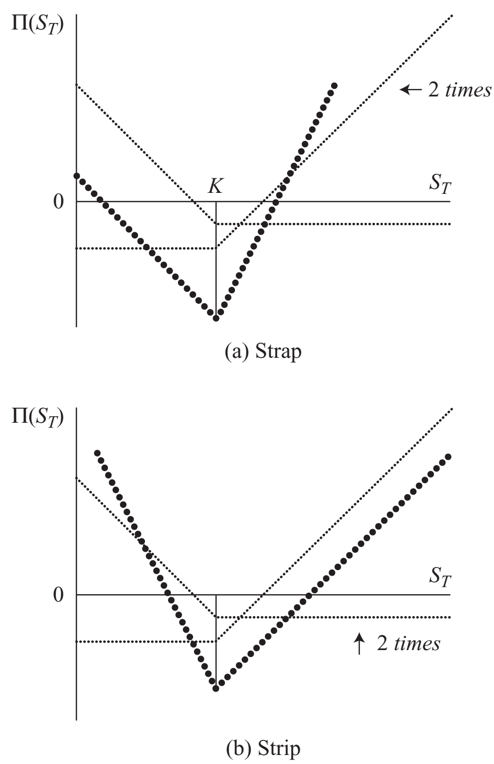
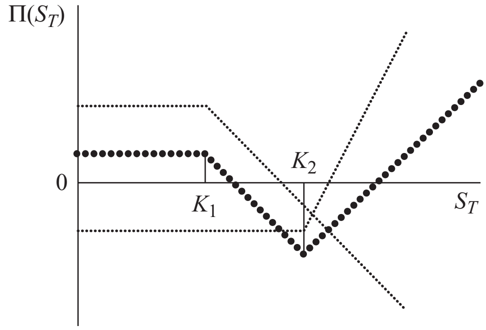
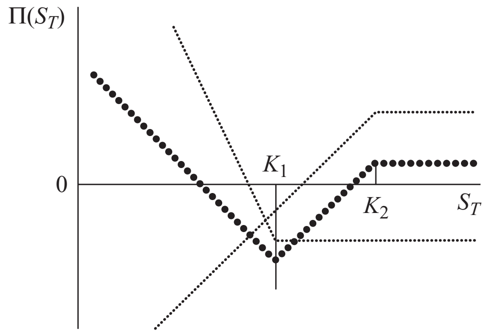


Exhibit 33.9 Profit/Loss Diagrams for Strap and Strip Positions



(a) Call Backspread



(b) Put Backspread

Exhibit 33.10 Profit/Loss Diagrams for Backspreads

RATIO SPREADS

Ratio spreads are composed of options at two different strike prices, with a ratio of written to purchased options other than 1:1. For example, consider the strategy of writing one call at a low strike and buying two calls at a higher strike. This is a 2:1 call backspread. Its payoff is like that of a straddle with a capped payoff on the downside. Similarly, a put backspread consists of written puts plus a greater number of purchased puts at a lower strike price. The put backspread has a payoff similar to a straddle with a capped payoff on the upside (see Exhibit 33.10). A Christmas tree is a variant of a ratio spread where a purchased call option at a low strike price is offset by written calls at two different higher strike prices.

BOX SPREADS

Consider the strategy of buying a call and writing a put with a low strike price K_L, and simultaneously writing a call and buying a put option at a higher strike price K_H. This position is simply a combination of a long synthetic forward with delivery price K_L and a short synthetic forward with delivery price K_H. The two forwards cancel each other out, and this position is equivalent to lending a small amount of money. The reverse strategy would be equivalent to borrowing a small amount of money. This strategy, known as a box spread, is one that an arbitrageur might take if the option prices are misaligned. Its payoff structure is summarized in the next table.

Payoff Table for a Box Spread

	$S < K_L$	$K_L < S < K_H$	$K_H < S$
Purchased K_L call	0	$S - K_L$	$S - K_L$
Written K_L put	$S - K_L$	0	0
Written K_H call	0	0	$K_H - S$
Purchased K_H put	$K_H - S$	$K_H - S$	0
Total	$K_H - K_L$	$K_H - K_L$	$K_H - K_L$

BUTTERFLIES, CONDORS, AND SEAGULLS

This section describes slightly more complicated positions that involve concurrent option positions at three or more strike prices.

A butterfly spread is an option position involving either three call options or three put options with evenly spaced strike prices. It is achieved by buying one contract at the low strike price K_L, buying one contract at the high strike price K_H, and writing two contracts at the medium strike price K_M.

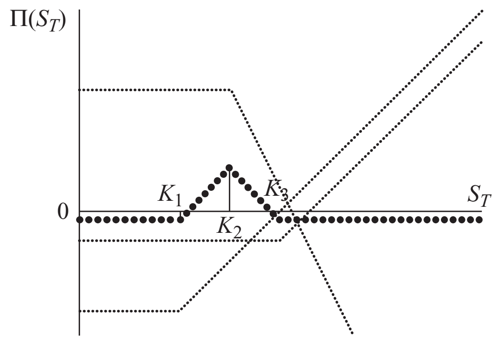
Payoff Table for a Butterfly Spread Using Calls

	$S < K_L$	$K_L < S < L_M$	$K_M < S < K_H$	$K_H < S$
Buy K_L call	0	$S - K_L$	$S - K_L$	$S - K_L$
Sell 2 K_M calls	0	0	$2(K_M - S)$	$2(K_M - S)$
Buy K_H call	0	0	0	$S - K_H$
Total	0	$S - K_L$	$2K_M - K_L - S$	0

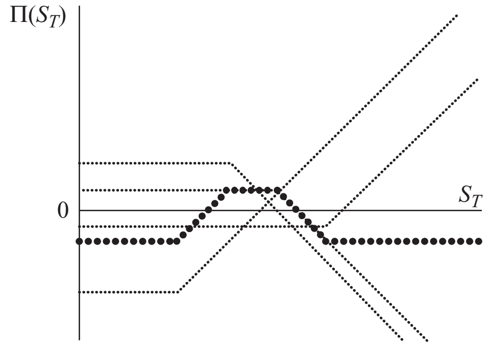
Payoff Table for a Butterfly Spread Using Puts

	$S < K_L$	$K_L < S < L_M$	$K_M < S < K_H$	$K_H < S$
Buy K_L put	$K_L - S$	0	0	0
Sell 2 K_M puts	$2(S - K_M)$	$2(S - K_M)$	0	0
Buy K_H put	$K_H - S$	$K_H - S$	$K_H - S$	0
Total	0	$S - K_L$	$2K_M - K_L - S$	0

If the prices of the three calls are represented by C_L, C_M, and C_H, it will cost $C_L + C_H - 2C_M$ to buy a butterfly spread. The owner of a butterfly spread receives a payoff only if the stock price ends up between the two strike prices. Like a written



(a) Butterfly



(b) Condor

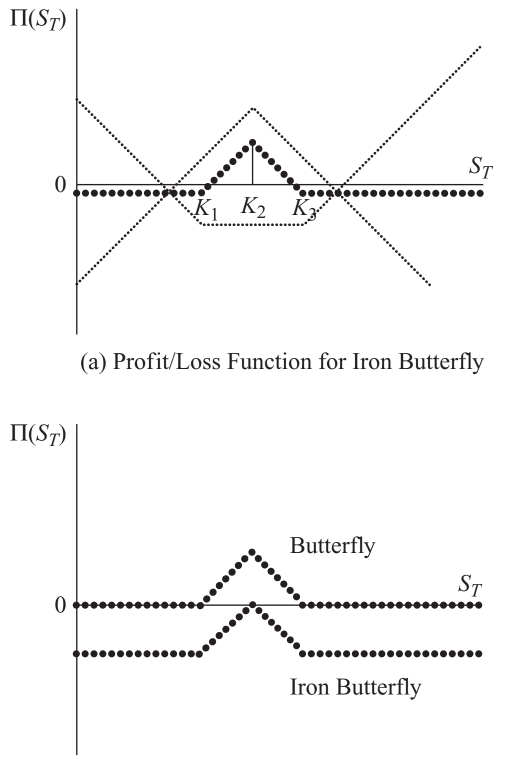
Exhibit 33.11 Profit/Loss Diagrams for Butterfly and Condor

straddle, a purchased butterfly spread may be viewed as a bet that the stock will end in a certain range. While a written straddle may generate large losses, the loss on a butterfly spread is limited to the initial investment. The payoff diagram is illustrated in Exhibit 33.11(a).

A *condor* is like a butterfly but involves options with four evenly spaced strike prices, the two written options having different strike prices. If there is an extra gap in the middle of the strike prices, some traders refer to the position as an albatross.

Payoff Table for a Condor Using Calls

	$S < K_1$	$K_1 < S < K_2$	$K_2 < S < K_3$	$K_3 < S < K_4$	$K_4 < S$
Buy K_1 call	0	$S - K_1$	$S - K_1$	$S - K_1$	$S - K_1$
Sell K_2 call	0	0	$K_2 - S$	$K_2 - S$	$K_2 - S$
Sell K_3 call	0	0	0	$K_3 - S$	$K_3 - S$
Buy K_4 call	0	0	0	0	$S - K_4$
Total	0	$S - K_1$	$K_2 - K_1$	$K_2 + K_3 - S$	0



(a) Profit/Loss Function for Iron Butterfly

(b) Payoff Function for Butterfly and Iron Butterfly

Exhibit 33.12 Iron Butterfly

Now consider the strategy of simultaneously (i) buying a strangle using "low" and "high" strike prices (K_L and K_H) and (ii) selling a straddle at a "medium" strike price (K_M) halfway between the low and high strikes. This strategy, called an iron butterfly, is equivalent to a butterfly spread plus borrowing. Exhibit 33.12 shows the profit/loss function for the iron butterfly and also how the payoff function for the iron butterfly differs from that of a standard butterfly spread.

Payoff Table for an Iron Butterfly

	$S < K_L$	$K_L < S < L_M$	$K_M < S < K_H$	$K_H < S$
Buy K_L, K_H strangle	$K_L - S$	0	0	$S - K_H$
Sell K_M straddle	$S - K_M$	$S - K_M$	$K_M - S$	$K_M - S$
Total	$K_L - K_M$	$S - K_M$	$K_M - S$	$K_M - K_H$

A long seagull is constructed by buying an out-of-the-money put and an out-of-the-money call while writing call option near-the-money. The name comes from the payoff diagram, which is reminiscent of a seagull, tilted in flight. The payoff profile of the long seagull roughly resembles a purchased call, but with an extra kink in the middle. The strategy might be used by an investor who is seeking the upside exposure and downside protection of a call option but would like the strategy to be self-financing. If the strike prices are chosen appropriately, the strategy should require no initial investment, as the initial income from the at-the-money call offsets the price of the two purchased options. A short seagull is constructed by buying an at-the-money put and financing it by selling an out-of-the-money call and an out-of-the-money put. The payoff roughly resembles a written call option with a kink in the middle.

TIME STRATEGIES

This chapter has focused on strategies involving various combinations of the underlying asset with calls and puts expiring at the same time. Other strategies involving options expiring at multiple maturities can be used to take a variety of positions that are sensitive not only to the level of the underlying stock but to the timing of the movements. For instance, in a bullish time spread or calendar spread, an investor may buy a long-term call and write a short-term call. This position is neutral over a shorter horizon but bullish over a longer horizon. Another variant is a diagonal spread, where the short-term and long-term options have different strike prices.

Investors also can use combinations of option strategies at multiple maturities to take positions that are sensitive to the timing of volatility. An investor who believes volatility will be low in November but high in December might write a November straddle and buy a December straddle. The payoff function for positions involving two maturities may be depicted in three dimensions, as illustrated in Exhibit 33.13.

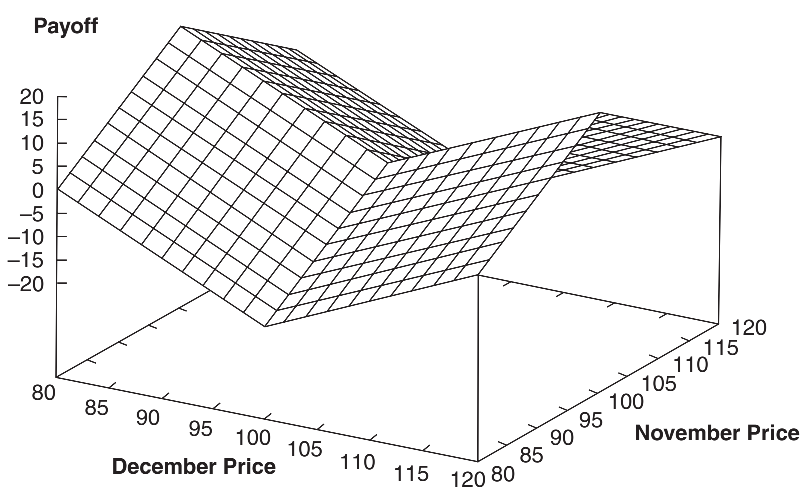


Exhibit 33.13 Long November Straddle and Short December Straddle

MULTI-ASSET STRATEGIES

The strategies outlined in this chapter all involve options on a single underlying asset. Of course, options on different underlying asset also can be combined as part of a single strategy. In futures markets, long/short spreads often are used to take exposure to the difference between two commodity prices (such as the difference between gasoline and crude oil). Options can be used to add nonlinear components to futures spreads, to create synthetic spreads, or to take offsetting volatility positions in correlated assets. For example, an investor who believes the volatility of the Nasdaq 100 index will increase relative to that of the the S&P 500 might want to buy a straddle on QQQQ and finance it by writing a straddle on SPX or SPY.

Combinations of options also can be used to take on exposure to the correlation between assets. For example, suppose an investor wishes to take a position reflecting a view that the correlation between the dollar/euro exchange rate and the euro/yen rate will be lower than the market's forecast. This can be accomplished by buying a straddle on the dollar/euro, buying a straddle on the euro/yen, and financing these purchased straddles by writing straddles on the dollar/yen.

Another strategy, called dispersion trading, involves offsetting a volatility position in an index option with volatility positions in the individual component options. For example, to bet on low correlation (high dispersion), an investor may write a straddle on the index and buy straddles on some or all of the component stocks.

ENDNOTES

1. This delta hedging can be accomplished by carefully offsetting the number of call and put options or by taking the appropriate offsetting position in the underlying asset. To remain delta neutral, the investor will have to update the hedge going forward as the underlying price evolves.
2. Chakravarty, Gulen, and Mayhew (2004) find evidence that some information about the level of stock prices is revealed first in the option market.
3. See, e.g., the Securities and Exchange Commission (2003).
4. See Cochrane and Saa-Requejo (2000).
5. For strategies involving purchased American-style options, the trader is free to ignore the early exercise feature and treat the option as if it were European. However, it is important to remember that a strategy that requires the trader to hold American-style options through their maturity dates may be suboptimal.
6. Alternatively, a profit/loss diagram may be constructed using the initial option price adjusted forward in time by the risk-free rate, eliminating the awkwardness of comparing cash flows occurring in two different periods.

REFERENCES

Chakravarty, S., H. Gulen, and S. Mayhew. 2004. "Competition, Market Structure and Bid-Ask Spreads in Stock Option Markets," *Journal of Finance* 57, no. 2: 931–958.

Cochrane, J., and J. Saa-Requejo. 2000. "Beyond Arbitrage: Good Deal Asset Price Bounds in Incomplete Markets," *Journal of Political Economy* 108: 79–199.

Merton, Robert C. 1973. "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4, no. 1: 141–183.

Ross, Stephen A. 1976. "Options and Efficiency," *Quarterly Journal of Economics* 90, no. 1: 75–89.

Securities and Exchange Commission. 2003. "SEC Interpretative Release Commission Guidance on Rule # 3b-3 and Married Put Transactions," 17 CFR Part 241, Release No. 34-48795 (November).

ABOUT THE AUTHOR

Stewart Mayhew is deputy chief economist at the U.S. Securities and Exchange Commission. He earned BS and MS degrees in economics from Brigham Young University and a PhD in finance from the Haas School of Business at University of California, Berkeley. Prior to joining the Commission, he served on the faculty of the Krannert School of Management at Purdue University and the Terry College of Business at the University of Georgia. He has published articles in the area of option market microstructure, option pricing, and volatility modeling, including numerous articles in the *Journal of Finance*, the *Review of Financial Studies*, the *Journal of Financial and Quantitative Analysis*, and the *Journal of Futures Markets*.

CHAPTER 34

The Use of Derivatives in Financial Engineering

Hedge Fund Applications

JOHN F. MARSHALL AND CARA M. MARSHALL
Lecturer of Economics/Finance at Queens College, City University of New York

INTRODUCTION

Over the past 15 years, the term *financial engineering* has come to be widely accepted as describing a way of thinking about and addressing financial issues in all areas of finance including corporate finance, asset management, investment finance, and financial institutions. Depending on their particular niche within the field, the term means different things to different people. But in general, financial engineering can accurately be described as "the development and creative application of financial technology to solve financial problems and exploit financial opportunities."[1] Financial engineering makes use of heavy-duty quantitative tools, the uses for which were once thought to be limited to physics and engineering. But also included in the financial engineer's tool kit is the entire spectrum of financial instruments. Perhaps most important among these financial instruments are derivatives.

There are many applications of financial engineering in corporate finance. Many actually are structured by financial engineers working for banks who pitch their solutions to corporate managements and boards.[2] Examples would include the use of currency swaps to obtain funding in currencies other than a corporation's domestic currency, perhaps to fund foreign operations, or funding in currencies other than the domestic currency and then swapping into the domestic currency because funding costs are cheaper in the foreign currency (even after allowing for the cost of the swap). Other applications include hedging the risks that corporations are exposed to including interest rate risk, credit risk, commodity price risk, and so forth. In other cases, by changing the method by which a corporate objective is achieved, or the products used to achieve it, there can be favorable accounting and/or tax implications.

Not only are many of the applications of derivatives forms of financial engineering, but the very design and analysis of new financial products, including new types of derivatives, are forms of financial engineering. In asset management, financial engineers structure new financial products to better appeal to the risk-reward appetites of investors. For example, there are investors who like to buy short-term notes but who want floating rates instead of fixed rates, or who want the performance of the note linked to the performance of some other asset class, such as equity (i.e., equity-linked notes), or gold or oil (i.e., commodity-linked notes), or to a specific credit that differs from the issuer's credit (i.e., credit-linked notes). They will even structure products that pay off based on which of several assets or asset classes performs better over a period of time.[3]

Other, relatively simple asset management applications include the use of asset swaps by fund managers to convert equity portfolios to synthetic fixed income portfolios and vice versa based on a temporary change in a strategic asset allocation plan. For example, a U.S. pension fund manager who held a diversified equity portfolio in early 2000, and who felt (correctly) the market was in a bubble, might have wanted to move from stocks to bonds for a while, say two years. But that would necessitate selling stock, buying bonds, holding the bonds for two years, then selling the bonds and buying stock (to return to the original strategic asset allocation). By simply entering into a properly structured two-year asset swap, the fund manager likely could have achieved exactly the same economic result with considerably less transaction cost.

The handiwork of financial engineers, whether they are using that title or not, shows up in all variety of risk management applications, most efforts at funding cost reduction, the development of more efficient trading platforms (such as electronic markets), and on and on.[4] While financial engineers wear many different hats, they share certain characteristics in how they approach problems. For example, one key characteristic of financial engineers is how they look at financial opportunities. Most often they see a financial opportunity as a bundle of risks. They look at each risk to determine whether they want to bear it. Then they systematically hedge away the risks they do not wish to bear. After factoring in the cost of hedging away the risks, they then ask, "Is the reward I expect to earn sufficient to justify the risks I am choosing to bear?"

Successful hedge fund managers are generally competent financial engineers, or employ competent financial engineers, who apply their talents to the world of investment finance.[5] For this reason, it is instructive to see how hedge fund managers might use derivatives in their strategies to extract what they like to call alpha. Two very good examples of this way of thinking about investment opportunities are convertible bond arbitrage and capital structure arbitrage. Both are strategies employed by quantitatively sophisticated hedge fund managers who specialize in these areas. In both strategies, derivatives play a key role.

CONVERTIBLE BOND ARBITRAGE

Today, over 90 percent of new convertible bond issuances (converts) sold in the United States and in Europe are sold directly to hedge funds—usually through private placements. These offerings tend to be large, averaging over a $500 million in the United States and even larger in Europe. Because of the size of the deals, a number of hedge funds will each take a portion of an issuance.

It is important to appreciate that most converts are sold by corporations that can accurately be described as weak credits. The reason for this is simple. A weak credit would have to pay a high coupon to sell its debt at par. Further, such corporations generally do not have any collateral to put up, thereby necessitating an even higher coupon. The issuer cannot afford such a high coupon. To get the investor to accept a lower coupon, the issuer offers the investor something else of value, namely a call option on the issuer's common stock.

Consider now a corporation that approaches the corporate finance desk (CFD) of an investment bank. The corporation indicates its desire to issue a convert in the amount of $500 million and a desire to sell it through a private placement. The CFD then refers the inquiry to the private placement desk (PPD). The PPD has a list of hedge funds that are in the market for convertible bonds for convertible bond arbitrage. So the PPD calls the hedge funds to ask them if they might be interested in the deal.

If a hedge fund responds that it may be interested, in which case it requires more detail, the PPD sends the hedge fund a nondisclosure agreement (NDA), which the hedge fund must execute and return before the PPD will send the hedge fund the term sheet and other specifics, including the name of the issuer. The NDA states that the hedge fund will not disclose any of the information being provided to any other party and that it will not act on this information if it chooses not to participate in the private placement.

The hedge fund sees the convert as a portfolio of instruments. Specifically, it sees a long position in a convert as a combination of a long position in a corporate bond and a long position in a call option on the same corporation's stock. Please see Exhibit 34.1.

The bond component can be viewed as a bundle of risks. Principal among these are interest rate risk and credit risk. There might be other risks if the convert happens to be callable or putable, but we will assume that it is not.

The hedge fund actually is only interested in the call option component of the convert and only then if the option can be acquired cheap. This brings us to a fundamental principle of option pricing. The value of an option, as long ago shown by Black and Scholes, is a function of five key variables that can be thought of as the value drivers of the option:

1. The current price of the underlying stock
2. The strike price of the option
3. The time remaining before the option expires (time to expiry)
4. The relevant interest rate for the time to expiry
5. The future volatility of the price of the underlying stock

We can see the current price of the stock, we can see the strike price of the option (at least in the types of options that Black and Scholes were modeling), we can

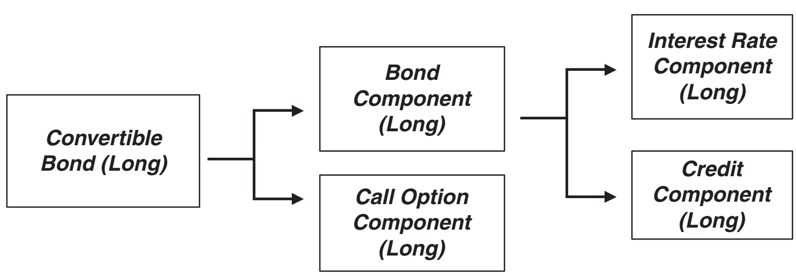


Exhibit 34.1 How a Hedge Fund Views a Convertible Bond

see the time to expiry, and we can see the relevant interest rate. But we cannot see the future volatility of the underlying stock. Thus, when you are trading options, what you are really trading is the one thing you cannot see. You are trading future volatility.

Volatility is measured as the standard deviation of the annual percentage change in the price of the underlying stock. The price of the option (i.e., the option premium) is directly related to the level of the perceived volatility. While the future volatility cannot be seen until after the future has become the past, it can nevertheless be inferred from the price of the option. That is, one can use an option pricing model to back out the volatility implied by the option's price. Such a volatility is called an implied volatility. Therefore, the option embedded in the convert is cheap if it can be bought at an artificially low implied volatility.

How do we determine if the implied volatility is artificially low? This is where the complexity arises. While the offering we have described is for $500 million, we will assume that our hedge fund would take only $100 million. Other hedge funds will take the other $400 million. For purposes of illustration, however, we will do the analysis in terms of $1,000 of par value rather than $100 million.

Suppose the term sheet says this:

> The convert has a life of 5 years and pays a coupon of 7.25 percent. [For simplicity, we will assume that this is paid once a year.]. The conversion ratio is 20 and does not change over the life of the convert. [That is, $1,000 of par value is convertible into 20 shares of common stock.] The option is only exercisable at the very end of its life [so it is of the European type]. The underlying stock does not pay a dividend and is not expected to pay one over the life of the convert.

The stock is currently trading at $40 a share. Because there is no clearinghouse for the embedded option, the relevant interest rate is not quite the risk-free rate.[6] We will use the five-year rate associated with a five-year interest rate swap as the relevant interest rate. (This is oversimplifying a bit, but it is a reasonable approximation for our purposes.)

The first thing the hedge fund manager would likely ask is: "What do I think the true future volatility of this company's stock is?" There are several ways to come up with an answer to that question. If there are long-dated options on this company's stock trading on one or more of the options exchanges, we could simply back out the implied volatility from those. Or we might look at the historic realized volatility and use that as an estimate of the stock's future volatility. This requires that we are willing to assume that the past volatility is indicative of future volatility. Or there might be comparable companies that do have options trading on them, and we could use an average of the implied volatilities from those options as a proxy for the future volatility of the subject company's stock. Suppose that one of these methods leads the hedge fund manager to conclude that the subject stock's volatility should be about 35 (i.e., 35 percent).

In order to determine if the option is cheap or rich, the hedge fund manager would need to determine four things:

1. The revenue that the convert will generate and the cost of funding the position.

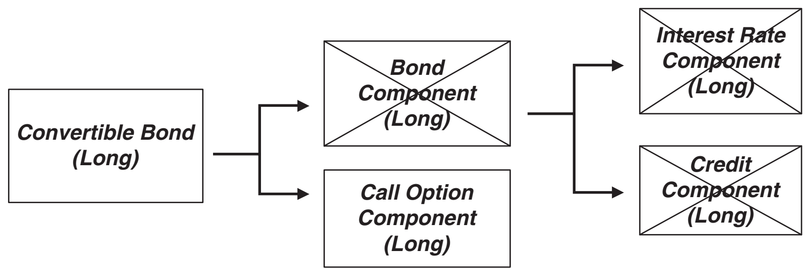


Exhibit 34.2 Stripping out Everything but the Option

2. The cost of stripping out everything but the option (meaning the manager has to get rid of the bond and get rid of the interest rate risk and the credit risk).
3. The implied price of the option.
4. From the price of the option the implied volatility. Please see Exhibit 34.2.

Let us continue with the example we have laid out. The hedge fund manager must first do the requisite analysis before deciding to do the trade or not. He begins with a series of phone calls. The first call is to the securities financing facility of the hedge fund's prime broker. This will be the source of most of the capital for the trade. That is, the hedge fund needs to borrow \$1,000 to buy the convert. Suppose that the prime broker says that it will lend the money at a rate of London Interbank Offered Rate (LIBOR) + 200 basis points (i.e., 2 percent). Please see Exhibit 34.3.

Based on this, the hedge fund manager sees that there is a fixed rate of 7.25 percent coming in (the coupon on the convert) and a floating rate of LIBOR + 2 percent going out.

This clearly exposes the hedge fund to considerable interest rate risk. That is, the hedge fund holds a fixed rate asset, but it is funded by a floating rate liability—never a good scenario. To address this problem, the hedge fund manager calls an interest rate swap dealer to price up a five-year plain vanilla interest rate swap in which the hedge fund would be the fixed rate payer and floating rate receiver. Suppose the swap dealer says the hedge fund would have to pay 6.10 percent against LIBOR flat. Please see Exhibit 34.4.

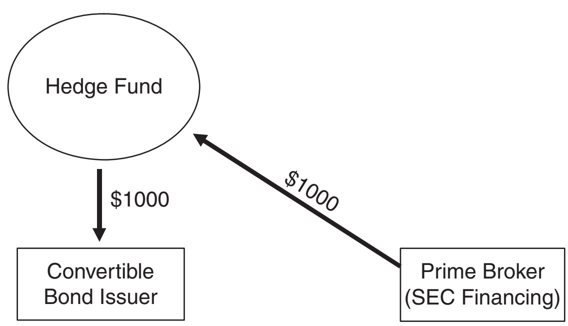


Exhibit 34.3 Borrowing to Buy the Convert

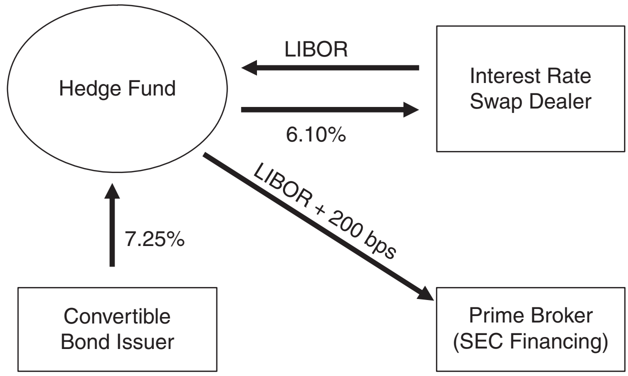


Exhibit 34.4 Hedging the Interest Rate Risk

The interest rate swap removes the interest rate risk in the sense that the LIBOR that is being received on the interest rate swap cancels the LIBOR portion of what is being paid to the prime broker. But the hedge fund manager still has the credit risk to deal with. To address this problem, he calls a credit default swap dealer and prices up a plain vanilla five year credit default swap (CDS). Suppose the CDS dealer says that it will sell credit protection for five years at an annual rate of 240 basis points. Please see Exhibit 34.5.

Since the interest rate swap eliminates the interest rate risk and the credit default swap eliminates the credit risk, the hedge fund manager has, for all intents and purposes, eliminated the bond component of the convert. He now needs to determine what the option cost. For purposes of this exercise, let us assume that all of the cash flows above are made once a year at year-end. The hedge fund manager will treat monies paid out as costs and monies received as reductions in cost. That is, the cost of the option = LIBOR + 2% + 6.10% + 2.40% − 7.25% − LIBOR = 3.25%. That is, the option appears to cost 3.25 percent of the convert's par value. Since the par value is $1,000, this is $32.50. However, this is not quite right for two reasons.

1. The option covers 20 shares of the stock (i.e., the conversion ratio is 20). Thus we have to divide by 20 in order to get the cost of the option per share covered. This is $1.625.

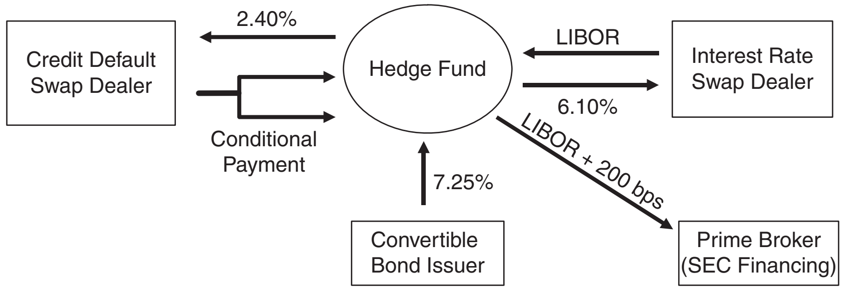


Exhibit 34.5 Hedging the Credit Risk

2. The hedge fund has to pay this each year for five years. So the true cost of the option, when viewed as a premium paid up front, is the sum of the present values of five annual payments of $1.625. Using the interest rate on the swap, 6.10 percent, as the discount rate, we get a present value of $6.83.

We can now know that the price the hedge fund is paying for the option is $6.83, the life of the option is five years, the option is of the European type and the underlying stock does not pay dividends, the stock is currently priced at $40 a share, and the strike price of the option is $50 per share. The strike price is $50 because the option is only exercisable at the very end of the convert's life at which point the hedge fund has a choice of taking $1,000 or taking 20 shares of stock. Thus, the hedge fund surrenders $50 for each share of stock it chooses to take. Finally, we will use the interest rate swap rate as a proxy for the five-year interest rate. Since we have all of the value drivers for the option except volatility, we can extract the implied volatility for the stock from a Black-Scholes model.

The result is an implied volatility of 15 (i.e., 15 percent).[7] Since the hedge fund manager believes that the volatility (vol) is actually closer to 35 and he can buy it through this rather complex process for 15, he would likely conclude that he can indeed get the option cheap. At this point, it would seem that the hedge fund manager should call everyone back and say, "Yes I will take the convert, yes I will take the funding, yes I will do the interest rate swap, and yes I will do the credit default swap"—that is, that he will pull the trigger on the trade. But we are not quite ready for that. The problem is that the hedge fund manager has managed to buy only the option cheap. To be arbitrage, the hedge fund manager must buy the option cheap and simultaneously sell it rich. That is, buy it in one market at a low volatility and sell it in another market at a higher volatility. So, before he pulls the trigger, he has to determine if, and how, he can sell it rich.

There are two possible ways to sell the option rich. The first, and the easier of the two, is to write call options on this same stock at the higher vol. This, however, is possible only if there are options on the stock trading on an options exchange (i.e., exchange traded, or ET) and there is enough depth and liquidity in these options to sell a sufficient number of them at the higher volatility. Alternatively, the options can be written in the over-the-counter (OTC) market, but the same liquidity concern applies. Please see Exhibit 34.6.

Unfortunately, it is quite likely that that will not be possible because either the options do not trade, or there is insufficient liquidity to sell the number of options the hedge fund manager needs to sell. Recall that the hedge fund manager is not really buying $1,000 of par value, but rather $100 million. Other hedge funds, which are using the same strategy, will be buying another $400 million. Collectively, the convert issuance is convertible into 10 million shares. That is substantial relative to the size of the options market.

That brings us to the second way to sell the option rich-by creating it synthetically. Or, more accurately, replicate its behavior synthetically through a process called delta hedging. The delta of an option is simply the change in the value of the option that would be caused by a change in the value of the underlying stock. The delta of a call will always be between zero and +1 (from the perspective of a long position). The more deeply a call is in-the-money, the closer its delta will be to 1. The more deeply a call is out-of-the-money, the closer its delta will be to zero.

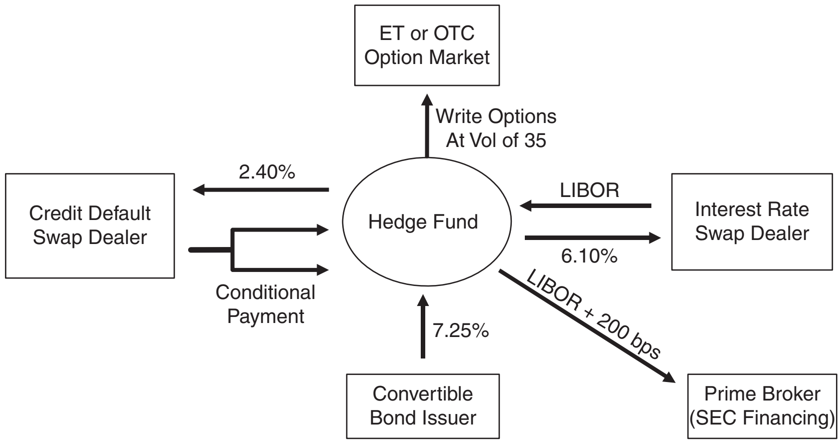


Exhibit 34.6 Selling the Options Rich

With the given set of value drivers—stock price is \$40, strike price is \$50, time to expiry is 5 years, interest rate is 6.10 percent, and the hedge fund manager believes the true volatility to be about 35—we can calculate all of the option's Greeks, of which delta is one. In this case, the option's delta turns out to be +0.69. That is, for each \$1 that the stock price changes, the value of the call will change by \$0.69 in the same direction. (This ignores the gamma effect.)

To delta hedge this option, all the hedge fund manager has to do is recognize that the delta of the stock is always 1.0. That is, the ratio of the change in the price of the stock to the change in the price of the stock is 1, because anything divided by itself is 1. Thus, the hedge fund manager needs to sell short 0.69 shares of the stock for each share covered by the option. For \$1,000 of par value, which is convertible into 20 shares, the delta hedge would require that the hedge fund manager go short 13.8 shares (i.e., 69 × 20). For \$100 million, the delta hedge would require that the hedge fund manger go short 1,380,000 shares. For the \$500 million of the entire offering, the hedge funds collectively will go short 6,900,000 shares of the stock.

Where does the hedge fund manager get the stock to sell short? The answer is from another specialized unit at his prime broker. This unit is called the securities lending facility. The securities lending facility will arrange to borrow the stock and lend it to the hedge fund. The proceeds from the short sale of the stock (plus a bit more as a cushion) will then be turned over to the prime broker to be held as collateral. The prime broker will invest this money in safe short-term vehicles earning a sum that will approximate LIBOR. The prime broker, however, will keep a little of this for its services and pay the rest out to the hedge fund. Let us suppose the prime broker keeps 25 basis points (bps), so the hedge fund gets LIBOR – 0.25%. Please see Exhibit 34.7.

The delta hedge in which our hedge fund manager borrowed and sold 1,380,000 shares would seem to solve the problem. But we have a few more issues to address.

1. The cost of borrowing stock (the 25 bps paid to the prime broker) increases the cost of the option we bought or, equivalently, decreases the price of the synthetic option we sold, and this must be factored in.

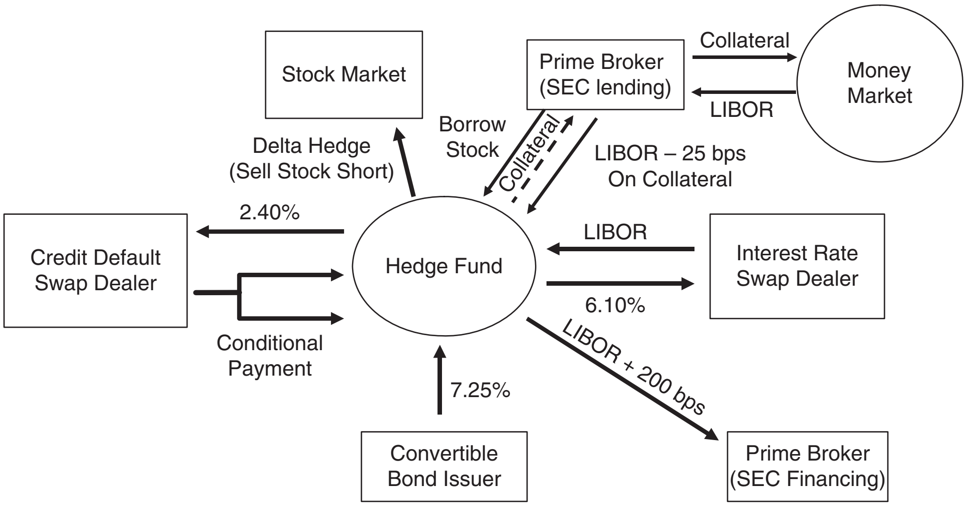


Exhibit 34.7 Delta Hedging

2. When this hedge fund and the other hedge funds that participate in this deal collectively sell short 6,900,000 shares of the issuer's stock, the price of the stock is undoubtedly going to be impacted. This will drive the price of the stock down with the effect that the delta will change (due to the option's gamma). This gamma effect needs to be taken into consideration when sizing the delta hedge.
3. An option's delta is not static. That is, the passage of time will cause the delta of the option to change. Fluctuations in the stock price and fluctuations in the level of interest rates will also cause the option's delta to change. This means that the hedge fund manager must periodically recompute the appropriate delta and adjust the size of the delta hedge. Not to do so would expose the hedge fund to risks it probably does not want to bear. This is why delta hedging is often described as dynamic delta hedging.

If the convertible bond arbitrage just described is executed properly and if the delta hedge is managed properly, the hedge fund will accrue the value of the difference between the volatility bought and the volatility sold as time passes. Of course, the transaction costs associated with periodically adjusting the delta hedge will eat into this profit to some degree. For this reason, the hedge fund manager will not do the trade unless he perceives the discrepancy between the implied volatility in the option embedded in the convert and the true volatility to be of sufficient magnitude.

The trade we have been discussing has been around for a number of years and is reasonably well understood. Some people consider it a type of volatility trading (which it is). Some people consider it a type of capital structure arbitrage (which it is). Some people consider it a type of relative value trade (which it is). And some people just like to think of it as a specific type of trade unto itself.

CAPITAL STRUCTURE ARBITRAGE

Capital structure arbitrage involves buying one component of a corporation's capital structure (such as its bonds) and simultaneously selling another component of the same corporation's capital structure (such as its stock). By this definition, it is easy to see why some people consider convertible bond arbitrage a form of capital structure arbitrage. As we saw, in convertible bond arbitrage, we bought a corporation's convertible bond and we simultaneously sold the same corporation's stock in an effort to delta hedge the embedded option. Nevertheless, when most people talk about capital structure arbitrage, they are not talking about convertible bond arbitrage.

Capital structure arbitrage rests on two key ideas. The first is conceptually simple but computationally tedious. We will not delve into the intricacies of the calculations. Rather, we will confine ourselves to the inherent logic.

All other things (i.e., interest rates, coupon payments, recovery rates in the event of default, etc.) being equal, the price of a corporate bond implies the probability that the bond will default over some period of time. This makes intuitive sense. If the price of the bond declines, it must mean that the market has concluded that the probability of default has gone up. Similarly, if the price of the bond rises, it must mean that the market has concluded that the probability of default has gone down.

As it turns out, if we have a number of different bonds from the same issuer, say a one-year bond, a two-year bond, a three-year bond, and so on, we can back out the entire term structure of default probabilities from the bonds' prices. That is, we can determine the probability of a default over the next year, the probability of a default over the second year, the probability of default over the third year, and so forth. The process is akin to the extraction of spot zero rates from Treasury bonds via the algorithm known as bootstrapping.

The second key idea is conceptually tricky and also computationally complex unless you are willing to make a number of simplifying assumptions. The idea originated with a paper by Robert Merton published in 1974. In this paper, Merton argued that a credit-risky corporate bond may be viewed as a portfolio consisting of a credit-risk-free bond and a short position in a put option on the company's assets. In 1993, Moody's Investor Services KMV unit (Moody's KMV) showed that Merton's argument implies that a company's common stock can be viewed as a long call on the company's assets. At that moment, the seeds of capital structure arbitrage were planted, but they needed almost another decade to take root.

Why can we think of a company's common stock as an option on the company's assets? The best way to answer this question is by way of an illustration. Suppose that the accounting values of a company's assets are consistent with the market values of the company's assets. Suppose further that the company has $45 of assets for each share of common stock outstanding and has $30 of bonds for each share of common stock outstanding. All creditors of the company are holders of the bonds and the bonds take the form of 10-year zero-coupon bonds. By the tautology of balance sheets, the book value (but not the market value) of the company's stock must be $15.

Let us assume that the 10-year zero-coupon rate of interest is 5.25 percent and that the company's stock is trading for $27.38.

Now suppose that the company suffers losses so that the value of the company's assets declines. This will also, of course, result in a corresponding decline in the book value of the company's equity. Now ask yourself this: Could the book value of the company's equity ever become negative? The answer, of course, in an *accounting* sense, is yes. If the value of the company's assets declines below $30 per share outstanding, the accounting book value of the company's stock becomes negative. But now ask a more important question: Can the *economic* book value of the company's stock ever become negative? For the economic book value to become negative, it would mean that the shareholders would willingly pay others to take their stock away from them. This is, of course, nonsense. Shareholders in a corporation have absolute limited liability. The company can simply declare bankruptcy, thereby surrendering the assets of the company to the creditors (i.e., the bondholders). Let us assume that the company would declare bankruptcy the moment the value of its assets declined to the book value of its bonds, or, equivalently, when the book value of its equity reaches zero. Please see Exhibit 34.8.

This economic book value is essentially the intrinsic value of a call option, which is readily apparent from the graphic relationship between the intrinsic value of the equity and the value of the assets. That is, whenever you see a hockey stick, you know you are looking at an option.

However, we said that the stock is not trading at its intrinsic value of $15. Rather, it is trading at its market price of $27.83. Why? The answer is time value. In addition to intrinsic value, options have time value. Time value represents the potential for the option (in this case, the stock is the option) to acquire additional intrinsic value prior to the option's expiry.

By thinking of the common stock as an option on the company's assets, it becomes clear that the value of the stock (when viewed as an option) is a function of the usual option value drivers: the value of the underlying assets, the strike price of the option, the time to option expiry, the rate of interest, and the volatility of the underlying assets. We know all these values except the volatility of the assets. But we also know the market price of the option (i.e., the stock), so we can extract

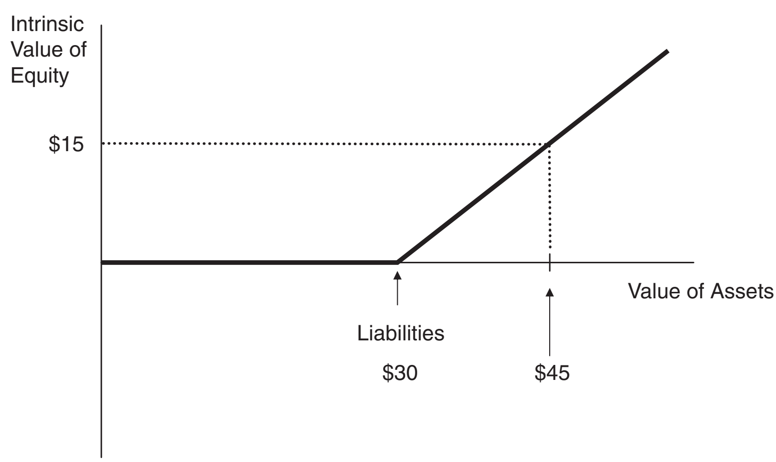


Exhibit 34.8 Economic Book Value of a Company's Stock

the implied volatility of the assets the way we would extract an implied volatility from any option premium.

In this case, the market price of the option (i.e., the stock) is \$27.83; the underlying asset is at \$45; the strike price is \$30; the time to expiry is the life of the bonds, which is 10 years; and the interest rate is 5.25 percent. Extracting the implied volatility of the assets, we get a volatility of 15 (i.e., an annual volatility of 15 percent).

We now ask ourselves another question: How far would the value of the assets have to decline for the company to declare bankruptcy, which would constitute default on its bonds? Well, we assumed that the company would declare bankruptcy the moment the value of its assets declined to the value of its liabilities, at which point its economic book value would be zero. That would require a decline of \$15 in the value of the assets from the current \$45 level to \$30. We can convert this to a percentage change (continuously compounded) by taking the natural log of the ratio of \$30/\$45.[8] The result is –40.5 percent. That is, if the company's asset value declines by 40.5 percent, the company will default.

Since a "vol" is simply a standard deviation, we can now ask the question: How many standard deviations would the value of the assets have to decline in order to result in default? This is just –40.5 percent divided by 15 percent, which is –2.70 standard deviations.

Assuming the percentage change in the value of the assets is approximately normally distributed, we can now just calculate the cumulative probability from negative infinity to a z score of –2.70. We get the result 0.35 percent. Please see Exhibit 34.9.

We can therefore conclude that there is 0.35 percent chance that this company will default on its bonds within a one-year period. We can extend this to get the probability of default over a two-year period, a three-year period, and so on to derive the entire term structure of default probabilities.

While many of our assumptions are more than a bit unrealistic, probably the most unrealistic is that the company would declare bankruptcy the moment the value of its assets declined to the par value of its liabilities (i.e., par value of its bonds). If that were the case, the typical recovery for bondholders following

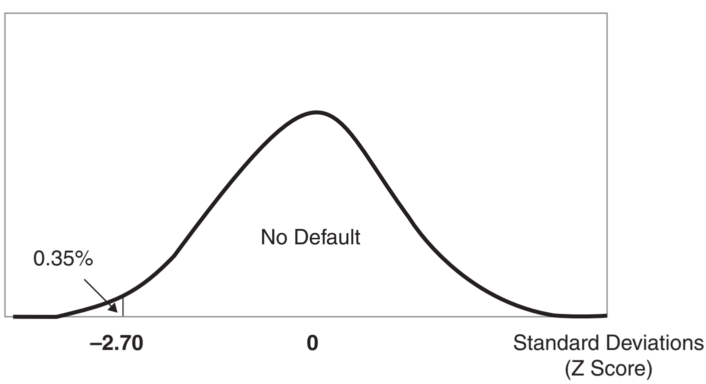


Exhibit 34.9 Cumulative Probability of Default

defaults would be 100 percent. In fact, even for senior secured bondholders, it averages only about 50 percent. Thus, companies generally do not declare bankruptcy until the accounting book value of equity gets well below zero (even though the economic book value of equity will never go below zero).

Thus, one must estimate at what point the company really would declare bankruptcy. The value of the assets at which a company really would declare bankruptcy is called the default barrier. The distance from the current value of the assets to the default barrier is called the distance to default and is most often measured in terms of standard deviations (vols). Please see Exhibit 34.10.

Now suppose that you extract the probability that a company will default on its bonds from the bond price, and you extract the probability the company will default on its bonds from the stock price, as we did. Now suppose that the bond price implies that the probability of default over the next year is 0.85 percent and the stock price implies that the probability of default is 0.35 percent. If markets were perfectly efficient (and if the underlying assumptions were completely realistic), the two implied probabilities of default should agree. But they do not. The higher probability of default implied by the corporation's bond price suggests that the bonds are cheap and the stock is rich. In this case, the hedge fund manager would buy the bonds and short the stock.

You might wonder what any of this has got to do with the derivatives that this book is about. The answer is in how we choose to execute the strategy. We can substitute credit default swaps for cash bonds, and we can substitute equity options for cash stock. Let us close this subject by exploring the logic of this just a bit.

If the bond price implies a relatively high probability of default, and if the CDS on the bond reflects a similar relatively high probability of default (relative to what is suggested by the stock price), we could simply enter into appropriate

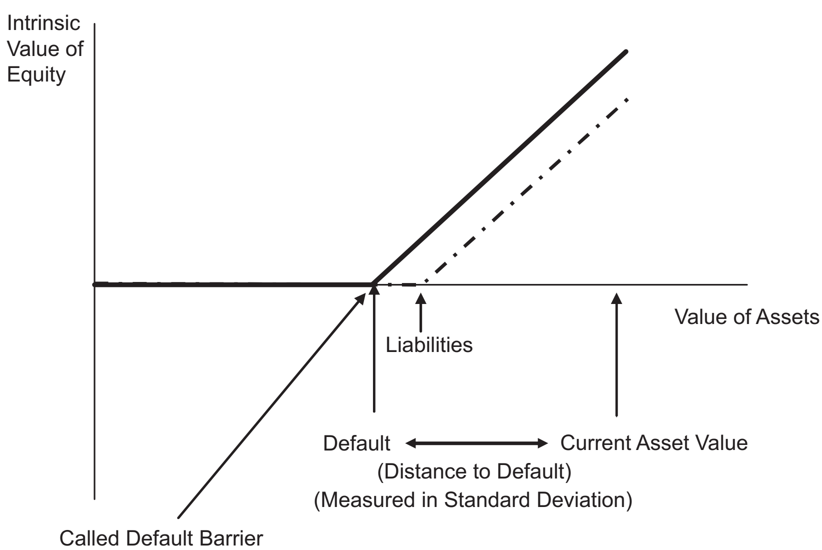


Exhibit 34.10 Distance to Default

credit default swaps as credit protection seller. However, because of the possibility that the CDS might not be efficiently priced, relative to the cash bonds, what we would really do is extract the probability of default on the bonds from the quoted spread on the CDS.

Now consider the stock. The volatility of the company's assets, as extracted by treating the stock as a call option on the company's assets, also implies the volatility of the company's stock. To see this, consider our earlier example in which we derived a volatility for the assets of 15. Now notice that there were $45 of assets for each $15 of equity. This means that the equity multiplier is three. (The equity multiplier is one of the leverage ratios and is defined as the ratio of the asset value to the book value of equity.) This implies that the volatility of the stock is 45. We would expect that options on the company's stock would have the same implied volatility (again if all our assumptions were realistic). But that will not necessarily be the case because the options might be mispriced relative to the stock.

For this reason, we can reverse the procedure. That is, look at at-the-money options on the stock and back out the implied volatility of the stock. Then, by dividing by the equity multiplier, we can infer the implied volatility of the assets. From the inferred volatility of the assets, we can obtain the probability of default on the bonds.

Suppose then that the spread on one-year CDS implies a probability of default that is higher than the probability of default implied by one-year at-the-money equity options. Then we would want to use CDS to sell credit protection and use equity options to sell volatility. Thus, we are playing both legs in a capital structure arbitrage with derivatives rather than with cash instruments.

If this logic is correct, then we would expect to find a high degree of correlation between the spread associated with the CDS on a company's bonds and the implied volatility associated with options on the company's stock, provided that they both have the same tenor. Repeated studies have indeed verified this to be the case.

There are many other examples we could offer of the use of derivatives in the hedge fund search for alpha opportunities, but the two provided in this chapter should be sufficient to make the case that many strategies employed by today's hedge fund managers would not be possible in the absence of derivatives and the mathematical logic that drives their pricing.

ENDNOTES

1. This was the original definition of financial engineering adopted by the International Association of Financial Engineers, the first professional society specifically for financial engineers, when it was launched in 1992. A similar definition was used by Marshall and Bansal (1992, p. 3).
2. For a good review of many of the applications of financial engineering in corporate finance, see Mason et al. (1995).
3. An excellent example of the latter is a product introduced by DWS Scudder (part of Deutsche Bank Group) called "Efficient Allocation Return Notes," or EARNS. This principal-protected zero-coupon structured product offers investors the opportunity to select multiple market indexes (e.g., S&P 500, Nikkei 225, and DJ Eurostock 50) with the payoff at maturity dominated by the best performing of the three indexes. The product incorporates a variety of exotic derivatives features including quantoes, rainbows, and Asians. See DWS Scudder (2007).

4. An excellent example of the role of financial engineers in developing new trading platforms was the development of the International Securities Exchange, which was the first all electronic options exchange in the United States. The efficiency of the exchange led to a dramatic reduction in the size of the bid-ask spread for equity options and forced the other options exchange to move away from their traditional floor-based platforms. See Chacko and Strick (2003).
5. A good example of this new thinking is James Simons, who runs the hugely successful hedge fund management firm known as Renaissance Technologies and who was awarded the 1996 Financial Engineer of Year Award by the International Association of Financial Engineers.
6. The interest rate plays two distinct roles in the pricing of an option. The risk-free rate (adjusted for the dividend yield) is used to determine the expected future value of the stock. Then an interest rate is used to discount the terminal value of the option to the present value of the option. The latter rate would be the risk-free rate only if there were no counterparty credit risk.
7. Any good option analytics software package should be capable of extracting this implied volatility. We used an internally developed package for this purpose.
8. When asset prices are lognormally distributed, the percentage change in their value will have a normal distribution when that percentage change is measured on the assumption of continuous compounding. Additionally, an asset's volatility (the vol) is routinely measured on the assumption of continuous compounding.

REFERENCES

Chacko, G., and E. P. Strick. 2003. *The International Securities Exchange: New Ground in the Options Markets*. Harvard Business School, Case # N9–203–063.

DWS Scudder. 2007. "Efficient Allocation Return Notes," filed with the Securities and Exchange Commission on August 31, 2007; SEC File 333–137902.

Marshall, J. F., and V. K. Bansal. 1992. *Financial Engineering: A Complete Guide to Financial Innovation*. New York: New York Institute of Finance.

Mason, S. P., R. Merton, A. Perold, and P. Tufano. 1995. *Cases in Financial Engineering: Applied Studies of Financial Innovation*. Upper Saddle River, NJ: Prentice-Hall.

Merton, R. C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance* 29, no. 2: 449–470.

Mitchell, M., L. H. Pedersen, and T. Pulvino. 2007. "Slow Moving Capital," *American Economic Association Proceedings* 97, no. 2 (May): 215–220.

ABOUT THE AUTHORS

John Marshall is the author of numerous books on financial products, markets, and analytics. He has also authored many articles in professional journals and is a frequently requested speaker for financial conferences. Dr. Marshall served on the faculty of St. John's University for 20 years before retiring in 2001. From 1991 to 1998, Dr. Marshall served as the executive director of the International Association of Financial Engineers, an organization that he cofounded. From 1994 to 1996, Dr. Marshall served as visiting professor of financial engineering at Polytechnic University, where he created the curriculum for the first Master of Science degree program in financial engineering under a grant from the Alfred P. Sloan Foundation. Dr. Marshall earned his undergraduate degree in biology/chemistry from Fordham University in 1973. He also holds an MBA in finance and an MA in quantitative economics. He was awarded his doctoral degree in financial economics from the State University of New York at Stony Brook in 1982 while also a dissertation fellow of the Center for the Study of Futures Markets at Columbia University.

Cara Marshall is lecturer of economics/finance at Queens College, City University of New York. Her research interests focus on derivatives and financial engineering as well as behavioral and experimental methods in finance. Her PhD dissertation examined the pricing of volatility on U.S. options exchanges. The first half of the study concerned the efficiency with which the market prices index volatility (as derivable from index options) relative to the individual volatilities of the index's components (as derivable from equity options). The second half of the study compared the implied volatilities of individual stocks derived in two different ways. Prior to academia, Dr. Marshall worked in Internet engineering, developing Web sites and an online platform for online course delivery. Over the years, she has served as a training consultant to several investment banks. In this role, she taught financial modeling to bank employees in New York, London, and Singapore.

CHAPTER 35

Hedge Funds and Financial Derivatives

TOM NOHEL
Associate Professor of Finance, Loyola University Chicago

INTRODUCTION

In a mere seven years hedge fund assets have grown from an estimated $323 billion in January 2001 to around $2 trillion today (see Exhibit 35.1).[1] This phenomenal growth has caught the eye of potential investors and regulators alike. The increased scrutiny has three main catalysts:

1. Once thought of as investment vehicles for the ultra rich, hedge funds now count pension funds and endowments among their biggest clients, potentially affecting many middle-income investors indirectly.
2. Hedge funds are important sources of liquidity in the derivative markets, particularly in the market for credit default swaps (CDS), where hedge funds constitute the bulk of non–interdealer trades and are estimated to participate in roughly 40 percent of all CDS.
3. Hedge funds often follow extremely complex trading strategies, showing far more sophistication than their mutual fund brethren and creating a certain mystique in financial circles.

In this chapter, I document the extent to which hedge funds utilize derivatives in their trading strategies, I highlight particular strategies in some detail, and I analyze the impact of hedge funds on the derivative markets and more broadly as well.

Commensurate with the tremendous growth in hedge fund assets, derivative markets have exploded in size, especially markets for credit derivatives (roughly 90 percent of which is the market for CDS), growing tenfold in the last *three* years to $51 trillion in notional value (up from $5 trillion; see Exhibit 35.2).[2] Given the confluence of a rapidly growing market of complex and hard-to-value securities, the dominant participation of the lightly regulated and quant-oriented hedge fund sector, and the still-fresh memories of the Russian debt crisis and the ensuing collapse of Long Term Capital Management and its orchestrated buyout by leading Wall Street investment banks, it is not surprising that former regulators, such as Gerald Corrigan, have sounded alarms about an impending crisis.[3]

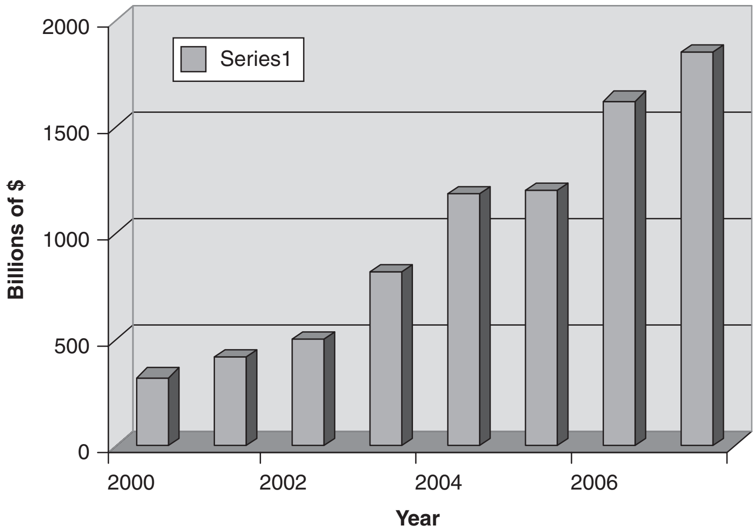


Exhibit 35.1 Hedge Fund Assets under Management

Given the amazing growth in hedge funds and derivatives alike, it might be tempting to characterize the surge in volumes in the derivatives markets as due primarily to trading by hedge funds. Although this may be true in certain sectors (e.g., credit derivatives), the overall volumes in the derivatives markets are far too large to be driven by the trading activity of hedge funds. As of the end of June 2007, the Bank of International Settlements (BIS) estimated the notional value of derivatives contracts outstanding at \$600+ trillion, with an estimated market value exceeding \$15 trillion and growing rapidly (see Exhibit 35.3), while the value of

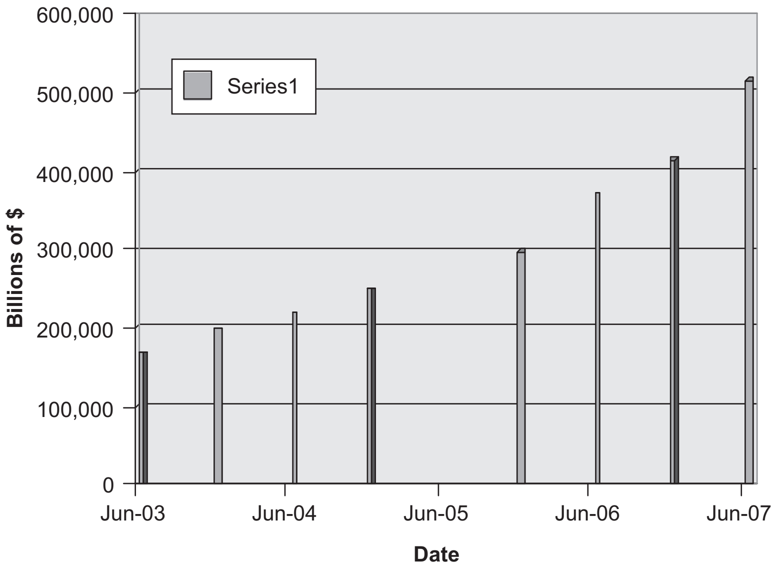


Exhibit 35.2 Notional Value of All Outstanding Over-the-Counter Derivative Contracts

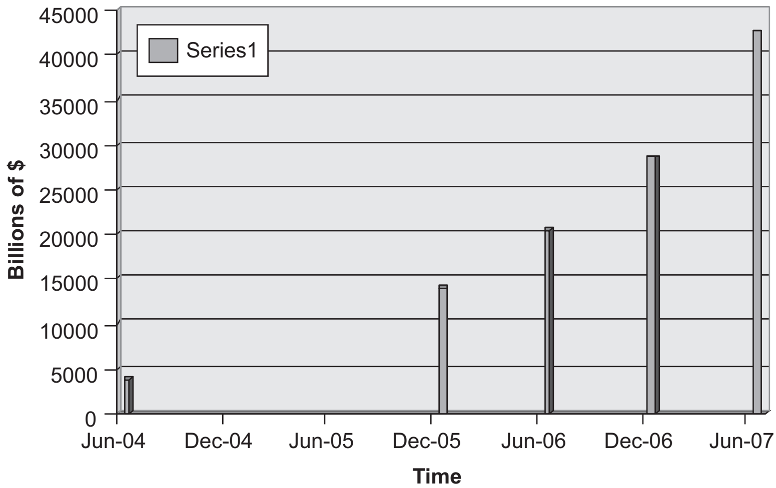


Exhibit 35.3 Notional Value of Outstanding Credit Default Swaps

assets under management at hedge funds is in the neighborhood of $2 trillion. Thus, even if we assume that hedge funds invest only in derivatives, and of course they do far more than that, it is impossible to explain the monumental growth in the derivatives markets as due primarily to the growth of hedge funds.

That being said, hedge funds are an important investor class in the derivative markets and a key source of liquidity, especially in the credit derivative markets, as noted earlier. In the credit derivative markets (dominated by credit default swaps), hedge funds represent 80 percent of *non–interdealer* trades and more than a third of all trades (according to the BIS). Moreover, credit derivative markets are the fastest-growing segment of the derivatives markets. Therefore, in a very real sense, the fortunes of the derivative markets, especially in the area of credit derivatives, are very much tied up in hedge funds. Though hedge funds are often credited with instigating the subprime crisis currently gripping markets worldwide,[4] they have in general fared better than most financial institutions.

I have five aims in this chapter.

1. I draw attention to the concurrent exponential growth in hedge fund assets under management and notional value of derivative positions, especially in the area of credit derivatives.
2. I seek to characterize the overall use of derivatives by hedge funds and its variation across funds, as well as the motivation behind these positions by giving a thorough summary of the Chen (2007) paper, which is the only comprehensive academic study of derivative usage by hedge funds.
3. I review the recent literature on the modeling of hedge fund risks and returns in light of their nonlinear trading strategies.
4. I briefly describe several popular hedge fund strategies that employ derivatives.
5. I give some anecdotal evidence on some unusual, and as a result lesser-known, derivatives trades by hedge funds. I close with some concluding remarks.

SURVEY OF DERIVATIVE USE BY HEDGE FUNDS

Given their embedded leverage, derivative contracts are very effective tools for managing risks. But the management of risk encompasses both the possibility of offsetting risks and the possibility of increasing exposure to risk in a particular direction (i.e., derivatives can be very effective tools for either risk reduction or risk taking).[5] We can easily find anecdotal evidence of hedge funds pursuing either or both of these strategies of derivative use. For example, the use of derivative contracts by a hedge fund interested in acquiring a voting block to be used in an upcoming corporate election without retaining the commensurate economic interest (i.e., empty voting) is clearly consistent with risk reduction.[6]

Alternatively, Zuckerman (2008) highlights the case of a hedge fund manager, John Paulson, who used derivative contracts in 2006 and 2007 to make bold directional bets on subprime mortgages and their derivatives. Specifically, Paulson's hedge funds sold short the riskiest tranches of collateralized debt obligations (CDOs) and then magnified that bet by buying default protection in the form of credit default swaps on similarly risky credits underlying the CDOs. Incidentally, Paulson's oldest credit hedge funds are purported to have posted returns of 590 percent in 2007, netting $16 billion of gains for investors. But which of these is the more common or widespread use for the typical hedge fund?

In the only comprehensive paper that I am aware of on the topic, Chen (2007) tries to answer this very question. Chen considers a number of aspects of derivative use by hedge funds, including the fraction of hedge funds that use derivatives overall broken down by primary strategy category, the types of instruments used (again broken down by primary strategy category), and the effect of derivative use on risk and performance, controlling for many factors that could affect derivative use by hedge funds. The paper has significant limitations due to the nature and availability of data on hedge funds.

The Chen (2007) paper uses hedge fund data provided by Tremont/Lipper known as the TASS database. To my knowledge, this is the only database that tracks derivative use by hedge funds. For example, Hedge Fund Research's HFR database does not track derivative use. However, the TASS database has several limitations, some of which are limitations of any hedge fund database. Since hedge funds are lightly regulated entities with minimal reporting requirements, there are few Securities and Exchange Commission (SEC) filings by hedge funds. For example, unlike mutual funds, hedge funds are not required to file a prospectus. Thus, any hedge fund data compiled for the purposes of conducting research is voluntary and subject to a reporting bias. Moreover, there does not exist a comprehensive database of hedge funds. Arguably TASS has the broadest coverage with more than 4,000 live funds reporting to the database and more than 2,000 dead funds tracked by the database, and it is survivorship bias free.

Another significant limitation is that TASS reports derivative usage as a 0/1 variable: Hedge funds either use derivatives or they do not. Then there are entries of 0/1 for each of a number of possible instruments. Therefore, funds that have derivative positions that number in the billions of dollars of notional values and funds whose position notionals are merely in the thousands of dollars are viewed identically. This also makes it difficult to draw conclusions on the exact use of the derivative positions employed since the magnitude of said positions remains a mystery. My intention is to summarize Chen's results in light of these limitations and then later focus on anecdotal details to get a clearer picture of how derivatives are used by hedge funds.

Chen (2007) reports that over the period 1994 to 2006, 73 percent of hedge funds used derivatives. There is considerable variation among fund styles, with over 90 percent of global macro and managed futures funds using derivatives while only about half of all market-neutral funds use derivatives. These numbers are five to six times the level of derivatives use by mutual funds.[7]

Why does this contrast in derivative use between mutual funds and hedge funds exist? While most mutual funds tend to be long-only funds that hold a diversified portfolio of long equity positions, many hedge funds follow complicated trading strategies that either involve derivative positions or create derivative-like payoffs (or both). These include such categories as risk arbitrage, convertible arbitrage, long-short equity, and more recently activist funds and credit-focused hedge funds. Although it is common for hedge funds to have a combination of long and short positions, mutual funds rarely do, but this is changing with the advent of 130/30 funds[8] and similar entities that involve some leverage.

The derivative use of hedge funds is consistent with these instruments helping to reduce transactions costs,[9] since hedge funds generally trade derivatives that are closely related to their core strategy. For example, while 79 percent of fixed income hedge funds use interest rate derivatives, only 9 percent use equity derivatives. Global macro funds tend to focus on currency derivatives (85 percent). Convertible arbitrage funds tend to use both equity and fixed income derivatives. Equity market-neutral funds are the least likely to use derivatives, which seems logical since *market* risk is minimal given the use of offsetting long and short positions (though we will come back to this point later). Finally, commodity derivatives are the least popular category of derivative with only 16 percent of hedge funds trading in these securities.

In terms of how various hedge fund characteristics affect hedge fund use, derivatives use was fairly stable over the sample period studied (1994–2006). Interestingly, hedge funds seem to favor certain types of derivatives. For example, users of equity futures tend to use futures in other categories if they trade derivatives in multiple categories. Derivative use is shown to be positively related to fund age, minimum initial investment, and incentive fees, whereas lockup provisions are likely to discourage derivatives use. Finally, managerial ownership and effective auditing of the fund are associated with a higher probability of a fund using derivatives. But how do derivatives effect hedge fund performance?

This is an important question that is likely to shed light on motives of hedge funds for using derivatives. Chen (2007) considers several measures of performance and several measures of risk in order to get a complete picture of how derivatives trading affects hedge fund performance and risk taking. He considers mean return, multifactor alpha, and the Sharpe ratio as performance measures and on standard deviation, market beta, idiosyncratic risk, downside risk, extreme event risk, skewness, kurtosis, coskewness, and cokurtosis as measures of risk.

There is considerable variation in the impact of derivative use on fund riskiness based on different measures of fund risk and different strategy categories. Only hedge funds categorized as directional show reductions in return standard deviation associated with derivative use. (Note: These tests lump hedge funds into four broad categories: Relative value, directional, fund-of-funds, and event driven.)[10] In contrast, strong reductions in market beta are associated with derivatives use in both the directional and fund-of-funds category. In fact, the betas of directional funds that eschew derivatives are nearly double those associated with funds that use derivatives, and in fund-of-funds, the comparable effect is even stronger. Interestingly, derivatives use generally is associated with higher levels of idiosyncratic risk. Downside risk is markedly lower for funds in the directional and fund-of-funds categories that use derivatives, and extreme risk is uniformly lower among funds that use derivatives, although only significantly lower among directional funds. But do these reductions in risk lead to performance improvement?

Results in Chen (2007) vary according to which performance measure is used. Not surprisingly, given the general reductions in systematic risk, mean return is lower among derivatives users relative to nonusers, especially at funds of hedge funds and directional hedge funds. Sharpe ratios are uniformly lower for derivatives users across all fund categories, but the Sharpe ratio is a notoriously bad performance measure to apply to hedge funds.[11] The best measure to consider is multifactor alpha. Here the strongest effect is seen among funds-of-funds, where risk-adjusted performance is significantly better among derivatives users. Other categories show no significant differences, suggesting in those cases that the risk reduction comes at the expense of sacrificing some performance. These figures are buttressed by tests based on a Carhart (1997) four-factor model and the model of Agarwal and Naik (2004) that includes option factors to capture nonlinearities, as well as the models of Fung and Hsieh (1997) and Harvey and Siddique (2000).

Chen (2007) also considers the effects of different types of derivatives. He categorizes derivatives as either linear (futures, forwards, swaps) or nonlinear (options). He finds that 59 percent of hedge funds use options while 51 percent of hedge funds use linear contracts. (Obviously a number of funds use both.) Users of futures appear to be more effective at reducing systematic risk while users of options are more effective at reducing extreme risk.

Finally, Chen (2007) considers the possibility of the impact of derivative use on risk-shifting behavior of hedge funds. There is considerable evidence from the mutual fund industry that managers adjust risk taking based on year-to-date performance. Specifically, Brown, Harlow, and Starks (1996) and Chevalier and Ellison (1997) find that managers who perform well in the first half of the year tend to get more conservative while managers who underperform their peers get more aggressive. Chen (2007) tests whether the same is true at hedge funds and whether derivatives use affects gaming behavior. He considers two measures of first-half performance: performance relative to primary strategy category peers as measured by raw returns and performance over the prior year as a measure of position relative to the high water mark.[12] For both measures of first-half performance, Chen (2007) finds that hedge funds that use derivatives are significantly less likely to employ a gaming strategy, consistent with a propensity to manage or reduce risk rather than to exploit the leverage inherent in derivative positions to ratchet up risk taking.

Chen (2007) concludes from his results that hedge funds' use of derivatives is consistent with the notion that these funds tend more to use derivatives as a tool to manage risks than as a tool to take risks. These conclusions are buttressed by three observations:

1. The most frequent users of derivatives (among hedge funds) are those that make directional bets on average.
2. Hedge funds that use derivatives are shown to have comparable (or lower) risk to their primary strategy peers that do not use derivatives.
3. Hedge funds tend to make use of derivatives based on similar underlying assets as their primary strategy category (e.g., equity funds use equity derivatives, bond funds use fixed income derivatives, etc.).

MODELING HEDGE FUND RISKS

With the explosion in hedge fund assets under management alluded to earlier has come increased scrutiny by regulators and academics alike. Regulators worry more and more about small investors who are indirect investors in hedge funds through pension funds and the like. Academics have paid a great deal of attention to modeling hedge fund returns and assessing hedge fund performance.

Early studies of hedge fund performance showed stupendous performance. For example, in analyzing performance in the area of risk arbitrage, Dukes, Frolich, and Ma (1992) and Jindra and Walkling (1999) found excess returns that exceeded 100 percent per annum. But these early studies of hedge fund performance used a linear factor structure to model returns as is commonly done in assessing mutual fund performance—for example, the Fama and French (1993) or Carhart (1997) structures. Soon researchers began to question the validity of this approach, given the complex trading strategies followed by hedge funds, including their propensity to use derivative securities to enhance leverage and (it is hoped) increase returns.

The earliest of these papers is Fung and Hsieh (1997), although it still proposes a linear factor structure for returns based on the asset class framework proposed by Sharpe (1992). But it was papers by Mitchell and Pulvino (2001), Fung and Hsieh (2001), and Agarwal and Naik (2004) that truly broke new ground by proposing that hedge fund returns have nonlinearities that cannot be captured by a linear factor structure.

Mitchell and Pulvino (2001) focus on risk arbitrage. Event-driven strategies such as this are popular with many hedge funds. The authors aren't able to replicate the stellar returns documented by Dukes et al. (1992) and Jindra and Walkling (1999) but are able to explain most of these excess returns as a combination of two components:

1. Transaction costs that are ignored in the earlier studies.
2. More important for our purposes here, compensation for significant event risk that is correlated with the overall market but in a nonconstant, and hence nonlinear, fashion.

The latter effect is consistent with the argument of Bhagat, Brickley, and Lowenstein (1987) who note that buying a target's shares after a deal is announced is analogous to holding a long share position plus a put option in the same shares.

In their tests, Mitchell and Pulvino (2001) find that the risk of deal failure is acute in down markets, creating significant nonlinearities in factor loadings.[13]

Similar conclusions are drawn by Fung and Hsieh (2001) who study so-called trend followers. They show that there are extraordinary risks involved in this strategy and that the payoff to the trend-following strategy is equivalent to an investment in a look-back straddle option spread.

Finally, Agarwal and Naik (2004) examine hedge fund returns over a broad range of strategies. Motivated by the Mitchell and Pulvino (2001) study as well as by an earlier paper of Glosten and Jagannathan (1994), Agarwal and Naik propose modeling hedge fund risk by estimating a multifactor model that includes both linear and nonlinear factors. They augment the four-factor model of Carhart (1997) by including at-the-money and out-of-the-money puts and calls on the Standard & Poor's (S&P) 500 index to capture the possibility of significant left-tail risk. The nonlinear factors are indeed priced, and the Agarwal and Naik approach has been widely adapted by academics who model hedge fund returns.

Agarwal and Naik (2004) attribute the nonlinear risk that they are picking up as the risk of a shift in volatility. They state: "If one can locate or construct an instrument whose payoff is directly related to volatility of financial markets, then it would be interesting to include it as an additional asset class factor" (p. 93). Identifying such a factor and estimating the market price of variance is precisely the goal in Bondarenko (2004). In his analysis, he considers both equity and nonequity hedge funds. Interestingly, he finds that the variance return, which is derived exclusively from the equity market, is a very important explanatory variable for equity and nonequity funds alike.

Bondarenko (2004) does not rely on any particular specification of the return process underlying options. Instead, he uses actual prices from S&P 500 index futures options. He finds that the price of variance risk is negative and economically very large. Given that hedge funds as a group are thought of as being sellers of volatility, Bondarenko considers the extent to which hedge fund returns can be explained as the funds being net short volatility. He finds that volatility risk can explain a significant portion of hedge fund returns and that a strategy of being short volatility would have paid off handsomely over the 13 years to 2003. These ideas can easily explain the findings of Mitchell and Pulvino (2001), Fung and Hsieh (2001), and Agarwal and Naik (2004), which largely focus on particular hedge fund strategies, such as risk arbitrage and trend following.

DESCRIPTION OF SOME POPULAR HEDGE FUND STRATEGIES

In this section I briefly describe several popular hedge fund trading strategies that tend to rely on positions in derivative securities.

Convertible Arbitrage

Convertible arbitrage is a popular hedge fund strategy that involves the simultaneous purchase of convertible securities and the short sale of the same issuer's common stock. As such the strategy is market neutral. Occasionally an interest rate hedge is applied in addition to hedging the equity risk by shorting the underlying shares. The idea behind the strategy is that there is a perception that convertibles often are priced inefficiently relative to the underlying stock. This is especially true when the hedge fund purchases the convertible in a private placement known as a PIPE (Private Investment in Public Equity).[14] As such, the equity option embedded in the convertible may be a source of cheap volatility, resulting in a source of value that convertible arbitrageurs can then exploit.

In convertible arbitrage, the number of shares sold short is based on maintaining a delta-neutral position. Thus, under *normal* market conditions, the arbitrageur expects the combined position to be insensitive to fluctuations in the price of the underlying stock. But the position may have to be revisited frequently and may require portfolio rebalancing (dynamic hedging). As with most market-neutral strategies, there are no guarantees that spreads will converge; in fact, they may widen under certain adverse conditions. For example, most convertible arbitrageurs suffered significant losses in 2005 when General Motors (GM) credit rating was downgraded and simultaneously Kirk Kerkorian was making an offer for GM shares. Since most convertible arbitrage funds were long GM debt and short GM equity, they were hurt on both sides. As with many hedge fund strategies, returns may be "supercharged" through the use of leverage, which can run as high as six to one. With that much leverage, an event like the Kerkorian episode in 2005 can be devastating.

Risk Arbitrage

Risk arbitrage, or merger arbitrage, is one of several trading strategies labeled as event driven and popular with many hedge funds. The risk arbitrageur attempts to capitalize on spreads between share prices of targeted companies and the value of bids for said targets. The exact strategy pursued depends on the financing of the proposed merger.

In a cash merger, an acquirer proposes to purchase the shares of the target for a given cash price. Prior to the completion of the deal, the stock of the target typically trades below the offer price. The arbitrageur buys the stock of the target and makes a gain if/when the deal goes through. In a stock-for-stock merger, the acquirer proposes to buy the target by exchanging its own stock for the stock of the target. An arbitrageur may then sell short the shares of the acquirer and buy the shares of the target, hoping to capture the spread between the bidder and target shares.[15] There is a risk that the deal will not go through or the closing will be delayed. Other problems may include inability to satisfy conditions of the merger, a failure to obtain shareholder approval, failure to receive regulatory clearances, or some other event that may change the target's or the acquirer's willingness to close the deal. Collars also can create difficulties for risk arbitrageurs, making the exact exchange ratio difficult to know in advance.

Global Macro

Global macro is a term used to classify the strategy of hedge funds that take positions in financial derivatives on the basis of forecasts and analysis about interest rate trends, movements in the general flow of funds, political changes, government policies, intergovernment relations, and other broad systemic factors or macroeconomic trends. George Soros famously employed a global macro strategy in his rather sizable bet against the British pound in 1992 at the time of the European Rate Mechanism. Soros purportedly earned over $1 billion from this trade and as a result is the best-known hedge fund manager in the global macro strategy sector. As noted earlier, global macro funds are among the biggest users of derivatives contracts.[16]

Another well-known strategy practiced by global macro managers that is meant to take advantage of the anemic Japanese economy is the carry trade, whereby hedge funds borrow large quantities of capital in Japanese yen and invest the proceeds in other currencies. Given that short-term interest rates in Japan are close to zero, hedge funds can borrow money and pay little interest and invest the proceeds in a country with relatively high real interest rates (a favorite of late is New Zealand) and amass large gains, provided of course that the relative values of the yen and the New Zealand dollar do not move in the wrong direction. As the Japanese economy began to show signs of life in 2007, many hedge funds lost considerable sums trying to unwind such carry trades.

Market Neutral/Relative Value

Market-neutral strategies (or the more broadly defined long-short equity strategy) are trading strategies that are widely used by hedge funds or proprietary traders. A fund goes long certain shares while shorting others in such a way that its portfolio has no net exposure to broad market moves. The goal is to profit from relative mispricings between long and short shares: going long those that are perceived to be underpriced while going short those that are perceived to be overpriced, while avoiding systematic risk.

A close cousin of the market-neutral strategy is the relative value strategy. Here the idea is to identify two baskets of securities that are essentially equivalent but yet have differing costs. The relative value fund would go long the basket that is inexpensive and short the basket that is more expensive. This strategy may involve positions in equity, fixed income, or derivative securities. This strategy was made famous by Long Term Capital Management prior to its high-profile collapse in 1998. Earlier, John Merriwether made substantial profits for Salomon Brothers using his famous on-the-run/off-the-run trade—a form of relative value trade. Trades were based on shorting on-the-run Treasuries (i.e., the most recently issued Treasuries) and using the proceeds to buy slightly shorter maturity (off-the-run) paper.

Volatility Trades

Volatility strategies trade volatility as an asset class, employing arbitrage, directional, market-neutral, or a mix of types of strategies, and include exposures that can be long, short, neutral, or variable to the direction of implied volatility. They can include both listed and unlisted instruments. Directional volatility strategies maintain exposure to the direction of implied volatility of a particular asset or, more generally, to the trend of implied volatility in broader asset classes. Arbitrage strategies employ an investment process designed to isolate opportunities between the price of multiple options or instruments containing implicit optionality. Volatility arbitrage positions typically maintain characteristic sensitivities to levels of implied and realized volatility, levels of interest rates, and the valuation of the issuer's equity, among other more general market and idiosyncratic sensitivities.

Correlation Trading

Correlation trading is a strategy in which the investor gets exposure to the average correlation of an index. The key to correlation trading is recognizing that the volatility of a portfolio is lower than the average volatility of the individual securities within the portfolio. Moreover, the less correlated the individual securities are, the lower the overall volatility of the entire portfolio. To buy correlation, investors can buy a portfolio of options on the index and sell a portfolio of options on the individual components of the index, or buy a variance swap on the index and sell the variance swaps on the individual components. Of course, to sell correlation, one simply does the reverse: Sell index options and buy options on individual components. This strategy can be implemented in any class of securities that has a liquid index traded on it.

Credit Hedge Funds

Among the biggest users of derivatives are credit hedge funds, although that moniker means different things to different participants. The HedgeFund Intelligence survey[17] categorizes credit hedge funds as part of its fixed income group. Others classify credit hedge funds as structured credit funds, or ones that trade only correlation, collateralized debt obligations (CDOs), and other forms of asset securitization. Still others see hedge funds that invest in equities as credit hedge funds—such as convertible arbitrage, capital arbitrage, or any other type of strategy that combines credit and equity. These trades come in many types, such as curve trades or cross-currency trades within a single credit (intracredit strategies), or intercredit strategies, such as trading Ford versus GM. There are also structured credit strategies (applying leverage over portfolios of credits) and relative value strategies, such as trading portfolios versus single names.

Some believe that hedge fund classifications are becoming more difficult to make because of the growing range of assets that they invest in. With the variety of asset types comes a growing complexity as well. On the credit side, that means anything from correlations to asset-backed securities, leveraged loans, and securitizations on all of them. The advent of more varieties of asset securitization means that there are always new forms of market imperfections for hedge fund managers to try to exploit. Structured credit trading has become more popular with hedge funds too, presumably because other markets and instruments, such as individual credits, have become less volatile, generating few opportunities for hedge funds, although some credit funds still use a few derivatives (e.g., those in high-yield and emerging markets).

There is apparently even a diverse spectrum of hedge funds even within correlation trading and structured credit. This can run the gamut from trading pure correlation on ITraxx index tranches to combining credit with correlation strategies to take directional views.[18] The development of ITraxx tranching has clearly had a significant impact on the type of credit risk plays hedge funds use. Another trend among hedge funds is the separation of default risk and spread risk. This is because iTraxx tranches have allowed credit players to trade different types of risk across the capital structure. This allows for new inefficiencies to be exploited, ultimately leading to more efficient pricing of credit risks.

With the advent of ITraxx, many different types of funds are able to use correlation products. These funds can use correlation products to express a leveraged view on the defaults of a portfolio of credits while staying neutralized against changes in market spreads. Among the most junior tranches there may be considerable default risk but little spread risk, while in higher-grade tranches the reverse is true. Added liquidity is also allowing hedge funds to do more with synthetic CDO tranches.[19] More specialized credit hedge funds have taken advantage of this by going long in the equity tranche and then protecting themselves in a more senior tranche referencing the same portfolio. The introduction of ITraxx products has facilitated that process by allowing hedge funds to go short or long in synthetic CDOs; previously they could only go long. For more details on these types of trades, see Horsewood (2005).

Hedge Fund Activism

A strategy that is all the rage among hedge funds is shareholder activism. In this strategy, hedge funds identify underperforming companies and purchase sizable stakes in these firms. The fund then pressures management to make changes that the hedge fund perceives to be value enhancing. Common examples are to divest or spin off certain assets, to subject top executives to more monitoring by opening up the board to more outsiders, or, in the extreme, to put the company up for sale. One might not expect this hedge fund strategy to actively employ the use of derivative securities, but I highlight a couple of examples next.

SOME UNUSUAL DERIVATIVES TRADES MADE BY HEDGE FUNDS

In this section, I give varying amounts of detail on some hedge fund trades that may be less familiar to many readers.

Empty Voting

With the surge in hedge fund activism has been a commensurate increase in the sophistication of the hedge funds involved, as highlighted by a recent example. In the summer of 2004, generic drug maker Mylan Laboratories decided to make a bid to acquire a controlling interest in King Pharmaceuticals. The deal was not well received by the markets, sending Mylan shares down over 15 percent on announcement. Sensing an opportunity, financier Carl Icahn purchased a 9 percent-plus stake in Mylan and immediately began to push Mylan's management to drop the takeover bid. Unbeknownst to Icahn, hedge fund Perry Capital (which had large positions in King) began building a large position in Mylan to counter Icahn and push the deal through. However, this position came with a twist: Perry had entered into sophisticated swap agreements with Goldman Sachs and Bear Stearns to completely hedge Perry's economic interest in Mylan.[20] Therefore, in spite of being Mylan's largest shareholder, Perry had no economic interest in Mylan. This is what Hu and Black (2006) refer to as empty voting.

Empty voting is a phenomenon that is purported to be widespread but is difficult to document. Allegations of empty voting swirl around several prominent takeover battles, including the long-running Hewlett-Packard/Compaq merger in 2002.[21] What makes Perry's role in the Mylan-King merger different is that the stake it had taken was so large that it was forced to reveal its positions in an SEC filing (unlike Citadel's rumored acquisition of a 4.4 percent stake, which did not breach the 5 percent barrier that would have required disclosure with the SEC). Carl Icahn later sued Perry for $1 billion, but both Icahn's suit and Perry's maneuvers were in vain because the deal fell apart for other reasons. The SEC is investigating Perry's actions.[22] Given that many economists are forecasting activism to be the focus of an increasing proportion of hedge fund investment, this is likely to be an important issue going forward. Moreover, a group of hedge funds working de facto in tandem can wield considerable influence over a shareholder vote without being detected if none of them individually has breached the 5 percent threshold.

Acquiring a Large Stake through Put Exercise

Among activist hedge funds, often secrecy is of the utmost importance. One method of acquiring shares that has little likelihood of being discovered (as long as one's total position does not exceed 5 percent of shares outstanding) involves the use of derivative securities (options) and is described next. Coghill Capital (a hedge fund) was able to acquire a sizable position in the stock of Energy Conversion Devices (Ticker: ENER) where it had intended to agitate for change. Coghill's technique to acquire a significant stake of 3,443,667 shares in ENER (representing a stake of 8.6 percent of outstanding shares) involved the *selling* of large numbers of put options. When the sold options were exercised (by others), Coghill managed to acquire a sizable stake without drawing a lot of attention to itself. Of course in this case, since the stake acquired surpassed 5 percent, Coghill was required to file a Form 13-d revealing its intentions (From Coghill's 13-D filing). This highlights another use of derivatives by activist hedge funds.

Toeholds via Contingent Contracts

In a recent article in the *New York Times*, Sorkin (2008) highlights the fact that activist hedge funds might now consider holding a position in a targeted firm through a contingent contract rather than through an actual share position as another means of avoiding detection of their intentions. This is taking the idea of empty voting and turning it on its head. In this instance, the hedge fund retains an economic interest but not a voting interest (opposite of the usual case).

Here is how it worked in a very recent case (attempted play for CNET by two hedge funds). Two hedge funds, Jana and Sandell Asset Management, acquired the equivalent of more than a fifth of outstanding CNET shares without anybody knowing. Given the existence of the 13D rule, which requires investors to disclose stakes of more than 5 percent, and the Hart-Scott-Rodino Act, which requires activist investors like Jana to disclose when they invest more than about $60 million, how could the hedge funds suddenly own the equivalent of 21 percent of CNet's shares? How could it even be legal?

Sorkin clarifies:

> *Hedge funds and other activists have quietly begun exploiting a loophole. Investors like Jana and others—including Carl C. Icahn, Nelson Peltz, and even the giant Canadian bank Toronto-Dominion—have begun entering into complex "swap" agreements with investment banks to get around the rules. The banks buy the shares on the investors' behalf, but technically never transfer full ownership. Think of it as a pseudo-off-balance-sheet deal. Technically, these investors say, they don't own the shares at all, just the "economics" of them. (p. 1)*

What investors recently have noticed is that the law requires them to disclose only when they control more than 5 percent of the vote. The law says nothing about controlling the economics in a company. So hedge funds use "swaps" to buy shares and strip out the voting power that comes with them. Here is Sorkin's simplified version of how it works:

> *An investor calls an investment bank and says, "Please buy 100 shares of company X. You can hold onto those shares in your name—and technically, you can do whatever you want with them. In six months, if the shares have gone up, you'll owe me the difference. If they go down, I'll owe you. And for all the cartwheels you're doing for me, I'll pay you a small fee." (p. 2)*

This is yet another novel use of derivatives at activist hedge funds.

Tax-Avoidance Strategies

Yet another hedge fund tactic, favored by hedge funds domiciled offshore (especially in the Cayman Islands), is to use swap contracts to avoid taxes on hefty dividend payments. In this strategy, the hedge fund looks to benefit from the equivalent of a long position in company shares without actually owning the shares outright. Instead, it has entered into swap agreements that effectively pass through capital gains and dividend returns. Because the hedge fund does not technically own the shares, the equivalent of the dividend payment stream is not taxable. According to Raghavan (2007, 2008), this strategy was being peddled to hedge funds by several prominent Wall Street investment banks, especially by Lehman, which made the strategy a centerpiece of its appeal to offshore hedge funds.

A Cayman-domiciled hedge fund client of an investment bank (e.g., Lehman) purchases a block of shares in a company with a hefty dividend (e.g., a real estate investment trust [REIT]) and then sells the shares to the investment client and simultaneously enters into a swap contract whose terms specify that capital gains and dividends from the underlying share position will flow to the hedge fund (and capital losses will be reimbursed by the hedge fund) in exchange for a small interest payment. The payment under the swap contract nets the income from the share position to zero for the investment bank. Thus only the investment bank's income under the swap contract is taxable. Raghavan (2008) states that investment banks are purported to have saved their hedge fund clients as much as $1 billion in taxes since 2000 using such trades.

Using Structures to Create Leverage

Another special trait of hedge funds is that they often operate with considerable leverage. Of course, one of the easiest ways to generate leverage is through the use of structures. Such structures may or may not involve the use of derivatives. An extreme example from Tett (2007) follows:

> *Consider a typical hedge fund that might be two-times levered. Assume that this fund is partly funded with fund-of-funds money which is itself three times levered. Moreover, assume that this money is invested in deeply subordinated tranches of collateralized debt obligations (CDOs) which are nine times levered. In this example, the fund-of-funds money is levered more than 50-to-1! In other words, this hedge fund has less than $20,000 of capital backing every $1,000,000 invested. A fall in the value of the CDO paper by more than 2% will completely wipe out the capital that is backing these positions. With US home prices down 7%–8% nationwide over the past year, it is no wonder that defaults are on the rise, and more likely than not there will be some high-profile blow-ups related to these issues in the months ahead. (p. 1)*

CONCLUSION

As can clearly be seen, the exponential growth in hedge fund assets and derivative notionals is not a complete coincidence. Hedge funds are big players in the derivative markets, particularly the credit derivatives markets. Though many participants fear the "hot money" of hedge fund investors that can leave a market even faster than it entered the market, hedge funds play a crucial role in the functioning of derivative markets, often taking positions that others are unwilling to take thereby adding much needed liquidity to these markets.

We have seen from the Chen (2007) paper that hedge funds' use of derivatives is widespread—much more so than their mutual fund brethren, but perhaps not as all-encompassing as many might be led to believe. It is important to keep in mind that it is not simply the use of derivative positions that create nonlinearities in hedge fund returns; in many cases strategies such as merger arbitrage, trend following, and convertible arbitrage lead to nonlinear returns, especially on the down side. In the extreme, according to Bondarenko (2004), many hedge funds make most of their money by shorting volatility risk and are appropriately compensated (e.g., the risk of deal failure in the case of merger arbitrage).

It is somewhat reassuring that Chen (2007) finds that most hedge funds, on average, use derivatives to hedge rather than make extensive directional bets. However, that is not to say that hedge funds could not benefit a lot from more extensive hedging, especially in the areas of insuring against volatility risk.

Finally, the various anecdotes toward the end of the chapter highlight some unusual hedge fund trading strategies that recently have come to light.

ENDNOTES

1. See Barclay Trading Group (2007).
2. According to the 2007 Bank of International Settlements triennial report.
3. See Platt (2005).
4. Note that it was the high-profile blow-up at two large hedge funds at Bear Stearns that led to the market calamity in August 2007 and added the term *subprime* to the lexicon of

many Americans. As a follow-up, subprime has just been voted the "Word of the Year" by the American Dialect Society.

5. We tend to think of risk management as leading to a reduction in risk, but a broader definition could include any action that impacts on the riskiness of an organization's positions.
6. See Hu and Black (2006). The concept of empty voting is discussed in more detail later in the chapter.
7. Several papers—Koski and Pontiff (1999); Deli and Varma (2002); and Almazan, Brown, Carlson, and Chapman (2004) have studied derivative use by mutual funds. There is near-unanimous agreement that mutual funds are minimal users of derivatives (10–15 percent use derivatives), in spite of more than 50 percent permitting derivatives use in their prospectuses.
8. In a 130/30 fund the idea is to 30 percent of base capital dedicated to short positions and then attain 30 percent additional market exposure on the long side. This effectively allows the portfolio manager to have an additional 60 percent exposure to risk than a traditional long-only fund.
9. See Deli and Varma (2002).
10. See Agarwal, Daniel, and Naik (2009).
11. See Getmansky, Lo, and Makarov (2004) and Dugan (2005).
12. A hedge fund's high-water mark is the benchmark that needs to be surpassed in order for a hedge fund manager to receive his or her performance bonus. Moreover, the high-water mark is cumulative, meaning that if a fund performs poorly in year t, it must first overcome that poor performance in order to receive a performance bonus in year $t+1$ and beyond. The typical performance bonus is paid as 20 percent of returns earned in excess of the high-water mark.
13. Mitchell and Pulvino (2001) show that betas of merger arbitrage strategies are close to zero in flat or rising markets but rise to around 0.5 in severe downturns, implying that the typical hedge fund specializing in merger arbitrage incurs substantial left tail risk.
14. According to Brophy, Ouimet, and Sialm (2007), hedge funds are the predominant investors in convertibles. They refer to hedge funds as "investors of last resort" in these products.
15. Of course, it is conceivable that these positions could be taken in the options markets rather than cash share markets. This would add leverage to the strategy due to the leverage inherent in the option contracts. This would naturally exacerbate the asymmetric impact of up and down market sentiment creating substantial left-tail risk.
16. See Chen (2007).
17. Available at www.hedgefundintelligence.com.
18. ITraxx is an index of liquid credit default swaps. The iTraxx suite of indices are owned, managed, compiled, and published by International Index Company, which also licenses market makers. CDS indices allow an investor to transfer credit risk in a more efficient manner than using groups of single CDS. They are standardized contracts and reference a fixed number of obligors with shared characteristics. Investors can be long or short the index, which is equivalent to being protection sellers or buyers.
19. A synthetic CDO is a CDO written off a portfolio of CDS instead of off a portfolio of actual credits. Investors in the synthetic CDO sell protection in the CDS market. Like standard CDOs, the credit risk in a synthetic CDO is divided into tranches where the junior tranches are funded (meaning there are funds held by the CDO to service payments that need to be made), while the most senior tranche(s) are unfunded. Proceeds from funded tranches typically are placed in a guaranteed investment contract that pays an interest rate slightly below the London Interbank Offered Rate (LIBOR).
20. Presumably Goldman and Bear then hedged their swap positions by shorting Mylan shares. Citadel was rumored to have made similar trades to acquire a voting interest of 4.4 percent.

21. Other examples include trades related to AXA's merger with MONY in 2004 and Salomon Lew's acquisition of a sizable stake in Australian firm Coles Myer for the purposes of advances his interest in a proxy fight. See Hu and Black (2006) for more details.
22. See Sorkin (2006).

REFERENCES

Agarwal, V., N. Daniel, and N. Naik. 2006. "Role of Managerial Discretion in Hedge Fund Performance," *Journal of Finance*, forthcoming.

Agarwal, V., and N. Naik. 2004. "Risks and Portfolio Decisions Involving Hedge Funds," *Review of Financial Studies* 17: 63–98.

Almazan, A., K. Brown, M. Carlson, and D. Chapman. 2004. "Why Constrain Your Mutual Fund Manager?", *Journal of Financial Economics* 73: 289–321.

Bank of International Settlements. 2007. Triennial Surveys on Positions in Global Over-the-Counter Derivatives Markets at End-June 2007. November.

Barclay Trading Group. 2007. *Alternative Investment Database*. BarclayHedge.

Bhagat, S., J. Brickley, and U. Lowenstein. 1987. "The Pricing Effects of Interfirm Cash Tender Offers," *Journal of Finance* 42: 965–986.

Bondarenko, O. 2004. "Market Price of Variance Risk and Performance of Hedge Funds," Working paper, University of Illinois at Chicago.

Brophy, D., P. P. Ouimet, and C. Sialm. 2009. "Hedge Funds as Investors of Last Resort?" *Review of Financial Studies*, 22 (February): 541–574.

Brown, K., W. Harlow, and L. Starks. 1996. "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance* 51: 85–110.

Carhart, M. 1997. "On Persistence in Mutual Fund Performance," *Journal of Finance* 52: 57–82.

Chen, Y. 2007. "Derivatives Use and Risk-Taking: Evidence from the Hedge Fund Industry," Working paper, Virginia Tech.

Chevalier, J., and G. Ellison. 1997. "Risk-Taking by Mutual Fund Managers as a Response to Incentives," *Journal of Political Economy* 105: 1167–1200.

Coghill Capital Management. 2007. Form 13-D filed with the Securities and Exchange Commission, February 22.

Deli, D., and R. Varma. 2002. "Contracting in the Investment Management Industry: Evidence from Mutual Funds," *Journal of Financial Economics* 63: 79–98.

Dugan, I. 2005. "Heard on the Street: Sharpe's Gauge Is Misused," *Wall Street Journal* 31 August.

Dukes, W., C. Frohlich, and C. Ma. 1992. "Risk Arbitrage in Tender Offers: Handsome Rewards—and Not for Insiders Only," *Journal of Portfolio Management* 18: 47–55.

Fama, E., and K. French. 1993. "Common Risk Factors in the Return on Bonds and Stocks," *Journal of Financial Economics* 33: 3–53.

Fung, W., and D. Hsich. 1997. "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds," *Review of Financial Studies* 10: 275–302.

Fung, W., and D. Hsieh. 2000. "Performance Characteristics of Hedge Funds and CTA Funds: Natural versus Spurious Biases," *Journal of Financial and Quantitative Analysis* 35: 291–307.

Fung, W., and D. Hsieh. 2001. "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers," *Review of Financial Studies* 14: 313–341.

Fung, W., and D. Hsieh. 2002. "Benchmarks of Hedge Fund Performance: Information Content and Measurement Biases," *Financial Analysts Journal* 58: 22–34.

Getmansky, M., Lo., A., and I. Makarov. 2004. "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns," Working paper #4288–03, MIT Sloan.

Glosten, L., and R. Jagannathan. 1994. "A Contingent Claim Approach to Performance Evaluation," *Journal of Empirical Finance* 1: 133–160.

Harvey, C., and A. Siddique. 2000. "Conditional Skewness in Asset Pricing Tests," *Journal of Finance* 55: 1263–1295.
Horsewood, R. 2005. "In with the Crowd: International Securitization Report," Tavakoli Structured Finance. Available at http://www.tavakolistructuredfinance.com/isr1.html.
Hu, H., and B. Black. 2006. "Empty Voting and Hidden (Morphable) Ownership: Taxonomy, Implications, and Reforms," *Business Lawyer* 61: 1011–1069.
Jagannathan, R., A. Malakhov, and D. Novoknov. 2006. "Do Hot Hands Persist Among Hedge Fund Managers? An Empirical Evaluation," NBER Working paper #12015.
Jindra, J., and R. Walkling. 1999. "Arbitrage Spreads and the Pricing of Proposed Acquisitions," Working paper, Ohio State University.
Koski, J. and J. Pontiff, 1999. "How are Derivatives Used: Evidence From Mutual Funds", *Journal of Finance* 54: 791–816.
Mitchell, M., and T. Pulvino. 2001. "Characteristics of Risk and Return in Risk Arbitrage," *Journal of Finance* 56: 2135–2175.
Platt, G. 2005. "Growth in Credit Derivatives Begins Attracting the Attention of Policymakers, Regulators," *Global Finance* (October). Available at http://www.allbusiness.com/accounting/debt/930941-1.html.
Raghavan, A. 2007. "Happy Returns: How Lehman Sold Plan to Side-Step Tax Man," *Wall Street Journal*, September 17.
Raghavan, A. 2008. "Wall Street Facing New Senate Probe," *Wall Street Journal*, January 15.
Sharpe, W. 1992. "Asset Allocation, Management Style, and Performance Measurement," *Journal of Portfolio Management* 18: 7–19.
Sorkin, A. R. 2006. "SEC Plans to Accuse Hedge Fund of Violations," *New York Times*, January 11.
Sorkin, A. R. 2008. "A Loop Hole Gets a Foot in the Door," *New York Times*, January 15.
Tett, G. 2007. "The Unease Bubbling in Today's Brave New Financial World," *Financial Times*, January 19.
Zuckerman, G. 2008. "Trader Made Billions on Subprime," *Wall Street Journal*, January 15.

ABOUT THE AUTHOR

Tom Nohel is an associate professor of finance at Loyola University. Professor Nohel has a PhD in finance from the University of Minnesota and a bachelor's and master's degree in applied mathematics at the University of Wisconsin and Northwestern University, respectively. He has published extensively in such journals as the *Journal of Banking and Finance*, the *Journal of Corporate Finance*, and the *Journal of Financial Economics*, among others. He has presented and discussed papers at numerous national and international conferences as well as at leading universities and related settings.

Professor Nohel's work is widely cited among academics worldwide, and has also been cited in popular and trade publications including Bloomberg, *Institutional Investor Magazine*, and the *Wall Street Journal*, among others. He has received awards for his papers and is currently partially funded by a grant from INQUIRE-UK as well as a grant from the Chicago Mercantile Exchange.

CHAPTER 36

Real Options and Applications in Corporate Finance

BETTY SIMKINS
Williams Companies Professor Business and Associate Professor of Finance, Oklahoma State University

KRIS KEMPER
Ph.D. Candidate, Oklahoma State University

> *One of the strongest arguments that can be made for real options is that it provides a framework that bridges the longstanding gap between strategy and finance.*
>
> —Triantis and Borison (2001)

> *It's not the strongest of the species that survive, nor the most intelligent, but those that are the most responsive to change.*
>
> —Charles R. Darwin (1809–1882)

INTRODUCTION

The globalization of world markets and advances in technology make it imperative for companies to adapt quickly to the changing marketplace, regardless if they are a local business or global corporation. As Triantis and Borison (2001) note, real options techniques can bridge the gap between strategy and finance. As the Darwin quote notes, adaptability is the key to survival and success. While Darwin was referring to living species, this concept is applicable to all businesses whether they are public, private, nonprofit, or governmental. Using a "real options" approach to address the challenges of a global marketplace leads to better managerial decision making. Coupled with traditional capital budgeting techniques, it leads to long-term survival and success of corporations.

What are real options? Real options are option-like opportunities, such as business decisions and flexibilities, where the underlying assets are real assets; hence the term *real options.* These opportunities are based on managerial flexibility and are commonly found in the operation of many corporate assets. Examples include: the choice to expand into new markets and product lines when growth opportunities are available in the marketplace; the option to wait until more information is available (such as to delay investment and operating decisions in response to uncertainty in the marketplace); the option to switch between inputs, outputs, or processes (such as a dual-fired power plant that operates on two types of fuel and can switch from one fuel to another); the option to abandon or temporarily shut down operations when losses are being occurred; and hybrid options.

It is important to note that the value of real options is dependent on the skills and expertise of management to "exercise" the real option at the optimal time. With strong managerial oversight, such options can add significant value to the business enterprise and can greatly benefit the strategic success of a firm. Likewise, with poor managerial oversight, these options have little or no value because the options are overlooked or not exercised optimally.

In the remainder of the chapter, we provide a brief history of real options, discuss the differences between financial and real options, describe the types of real options together with examples from the energy industry, list ways real options are valued, and conclude.

A BRIEF HISTORY OF REAL OPTIONS

Until 1973, when Fisher Black and Myron Scholes published their path-breaking paper on pricing options, options valuation was not well developed because there was no well recognized or accurate method. Black and Scholes's work, which provided the first closed form solution to options valuation, laid the foundation for the birth of real options as a formal area of study.

While "real option-type" business opportunities have been around for centuries, the official coining of the words *real options* was by Stewart Myers (1977) of MIT when he observed that any capital budgeting decision was in fact a series of embedded "real options" on additional investments. Myers states: "Many corporate assets, particularly growth opportunities, can be viewed as call options. The value of such 'real options' depends on discretionary future investment by the firm" (p. 147). Prior to this period, real option decisions were based on intuition, ad hoc techniques, or rules of thumb with little quantitative theory behind the decisions. Myers formally introduced the topic as a new area of financial research. For this reason, he is viewed as the father of real options.

Since the 1970s, many articles and books have been written on the subject of real options, and continue to be written. Researchers first applied real options to valuation of oil and gas exploration rights. Oil and gas companies realized that reserves had "real option" value, even if the costs of developing were unprofitable at the present time, due to the high historical volatility of oil prices. These valuation techniques subsequently led to applications in other industries.

Real options are now recognized as a way of thinking, as a part of the capital budgeting process, as an analytical tool incorporated into software packages such as Crystal Ball, and as an integral component of corporate strategy and decision making.[1] However, the practice of real option analysis by businesses is still limited and is used most frequently in industries such as energy (oil, gas, and power) and other natural resources, aviation, transportation, pharmaceuticals, high tech, among others. There is a robust literature on applying real options to capital budgeting decisions in corporate finance.[2] Many businesses do not know about the process/technique/concept.[3]

DISTINCTION BETWEEN FINANCIAL OPTIONS AND REAL OPTIONS

Before discussing the types of real options, it is important to understand the distinction between financial and real options. Financial options give the holder the right, but not the obligation, to engage in a financial transaction at a predetermined price and date. More specifically, a *call option* gives the holder the right to buy the underlying financial asset (such as a share of stock) at some predetermined, or *exercise*, price. A *put option* gives the holder the right to sell the underlying asset at its *exercise* price. Conversely, real options give the holder the ability to acquire an asset using a call option, to divest using a put option, or to "exercise" other flexibilities on real assets. These options give a manager the ability to be flexible and can be viewed as strategic options (i.e., strategic opportunities that arise from owning real assets). In other words, financial options give the holder a right to purchase (sell) a financial asset while real options give the holder the right to purchase (sell) an actual physical or knowledge-based resource. Furthermore, financial options have exact exercise prices and dates, while real options are a function of the resources involved, the owner, and the environment.

To summarize:

Financial Options	**Real Options**
Right to purchase/sell a financial asset	Right to purchase/sell physical resource
Exact exercise date	Exercise date not specified by contract
Exact exercise price	Exercise price not specified by contract
Value does not depend on who owns it	Value generally depends on who owns it

TYPES OF REAL OPTIONS AND EXAMPLES IN THE ENERGY INDUSTRY

Real options have many useful applications in the energy industry. The existence of derivative securities, such as futures, forwards, and options on oil, petroleum products, natural gas, and power, makes it possible to more fully exploit the flexibility of energy assets, such as oil and gas reserves, natural gas storage facilities, pipelines, fertilizer plants, power plants, and alternative energy projects. (See Chapter 9.) Real option applications in these areas permit the optimization and valuation of the flexibilities embedded in the operation of these energy assets by their owners. Furthermore, the volatility of energy prices adds value to real options applications in the energy industry.

When considering a project, the traditional net present value (NPV) formula is modified to include the value of the real option.[4] In other words, the value of a project with real option features is equal to:

$$\text{Value of project with real options} = \text{NPV} + \text{Real option value} \tag{36.1}$$

Common types of real options and examples in the energy industry are described next. Exhibit 36.1 lists common types of real options together with examples and references for additional information. These types of real options are not mutually exclusive because real assets can have multiple option type opportunities embedded in them.

Exhibit 36.1 Common Types of Real Options

Types of Real Options	Option Type	Examples in the Energy Industry or Related	Selected References
Option to expand	Call	Southwest Airlines Blended Winglets; Replace oil or gas reserves by exploration or acquisition; Invest in new data acquisition (seismic data and logs, etc.) for energy exploration	Amram and Kulatilaka (1999); Copeland and Antikarov (2001); Damodaran (2000, 2008); Martin, Rogers, and Simkins (2004); Titman and Martin (2008
Option to wait	Call or put	Explore for oil or gas onshore or offshore; valuing PUD reserves; valuing oil sands; carbon capture investments	Damodaran (2000, 2008); Dixit and Pindyck (1994); Brennan and Schwartz (1985); McCormack and Sick (2001); Ronn (2005a);
Option to switch (inputs, outputs, or processes)	Call or put	Valuing a natural gas power plant; valuing natural gas storage facilities; flexibility of inputs with dual-fired power plants	Amram and Kulatilaka (1999); De Jong and Walet (2005); Kasanen and Trigeorgis (1993); Kulatilaka (1993); Miller and Waller (2003); Ronn (2005b); Thompson, Davison, and Rasmussen (2007); Triantis (2000); Trigeorgis (1993b)
Option to abandon or temporarily shut down	Put	Production of fertilizer; stripper well abandonment decision	Damodaran (2000, 2008); Ronn (2005a); Titman and Martin (2008);
Hybrid real options	Call or put and others	Game theory applications in competitive game situations	Dias (2005); Grenadier (2000); Smit and Ankum (1993)

Option to Expand

The option to expand allows a company to enter into a new market or project with caution or to move more aggressively if conditions warrant such action. In this scenario, the firm (holder of the option) has a call option on additional capacity. If conditions are favorable, the call option will be exercised and additional capital will be invested in the project.

As another application, a firm may have an option to extend the life of an asset at the end of a project. In this situation, the firm has a call option on the asset's future value. In this case, an asset anticipated as being obsolete at the end of a project instead has remaining economic life. The firm has a call option on the asset's future value and may exercise its right to extend the project.

Example: Tertiary Recovery Techniques in Oil and Gas Production
A common application to extend a project's life in the oil and gas industry (an option to expand recoverable reserves) is to use tertiary recovery techniques. Tertiary recovery in oil and gas production begins when secondary oil recovery is not enough to continue adequate production, but only when the oil can still be extracted profitably.[5] Hence, this potential opportunity can be viewed as a call option in which the option premium depends on the cost of the extraction method and the payoff is based on the current price of crude oil. When prices are high, the option is exercised and previously unprofitable wells are brought back into production.

Option to Wait

Instead of deciding today to invest in a project based on NPV analysis, a firm can wait and learn more before investing. This is also known as the timing option. In this situation, the firm has a call option on the value of the project. For example, this type of real option would be valuable when a firm is considering entering into a new line of business, one in which there is not much of a track record. A firm can opt to delay its decision and see if demand is high. If demand is high, the firm can exercise its call option and enter the market. However, if demand is low, the firm can avoid this market altogether without having made any financial investment.

Example: Valuing Petroleum Undeveloped Reserves as Real Options
The market value of oil and gas reserves, especially petroleum undeveloped reserves (PUDs), usually have a greater value than the value calculated using the traditional NPV approach. This is due to the real option opportunities embedded in such assets. Exhibit 36.2 illustrates this concept. Consider PUDs that have a NPV of \$100 million if developed now at a cost of \$100 million. According to NPV analysis, the value is zero. However, these reserves have positive economic value because the owner (or lease holder) of these reserves has years to wait before losing the right to develop. Hence, these assets have real option value from the flexibility to expand in the future. This option value can be viewed as a finite-lived American call option, with time to expiration established by ownership of the mineral rights or lease agreement. The real option value is the value of the right (but not

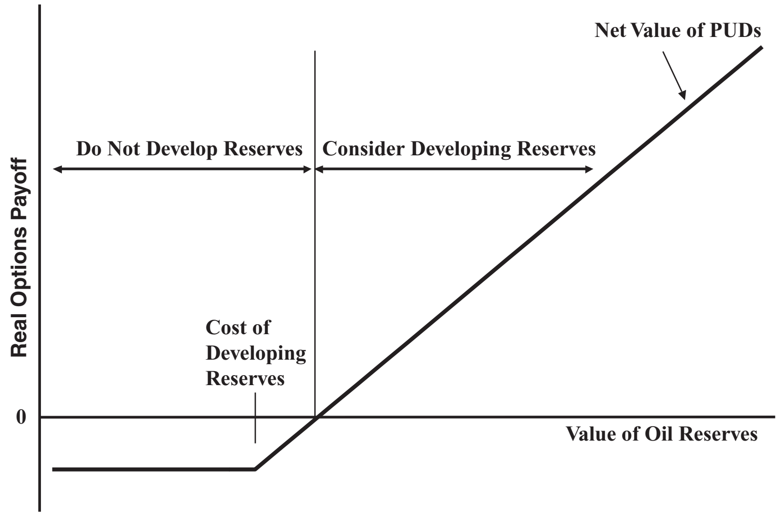


Exhibit 36.2 Value of Petroleum Undeveloped Reserves (PUDs) as a Real Option

obligation) to invest in the PUD during the life of the lease. The value can be modeled as following a geometric Brownian motion, among other techniques.[6]

Option to Vary Production Inputs, Outputs, or Processes

Standard NPV analysis ignores the value of flexibility in projects such as the option to vary production inputs, outputs, or processes. Such flexibilities can greatly enhance the economic value of a project. With flexibility in inputs and outputs, the company has the option in each time period to convert the lowest-cost input into the highest-value output (if there is more than one choice for the input or output). In essence, any company that builds a production system so that it can change easily from one input to another or from one product to another has built a real option component into the asset. The company only needs to exercise the real options in these flexible projects if it is profitable to do so.

With the values of inputs and output commodities fluctuating highly (i.e., high price volatility), the profitability can vary widely from high profits to large losses, especially if input and output prices are not highly correlated. The more volatile this profitability spread, the higher the value of this option and the greater the possible profits. This can be summarized: The more volatile the relationship between the prices of inputs and outputs, the greater the difference between the standard NPV and the true NPV (if inherent real options are exercised and included in the true NPV value).

The next examples describe applications in the energy industry. It is interesting to note that equity prices have exhibited correlations with real options valuation. For example, Dawson and Considine (2003) show that the stock price of an electricity generating company is significantly correlated with the volatility of spark spreads.[7] (Note: The spark spread represents the theoretical margin for a power plant. If the price of electricity (the output) is higher than the price of fuel (such as natural gas) for a power plant, then the spark spread is positive. Likewise, if the price of electricity is less than the cost of the fuel, the spread is negative.) The real options illustration described next provides more information on this topic.

Example: Valuing a Natural Gas Power Plant: The Tolling Arrangement

The tolling arrangement is an innovative structured deal that has been used by the power industry since the late 1990s. Exhibit 36.3 provides a simple illustration of the tolling concept. Tolling arrangements are the practice of operating a natural gas power plant for a fee under a contract to another party (the real options holder) who provides the input (natural gas fuel) and sells the output (the electric power produced). In other words, the real option holder takes on the market risk by supplying the natural gas used and selling the power produced and the counterparty owns and operates the power plant with its own employees. This concept has allowed market participants to capitalize on new market opportunities in the deregulation of the power industry. The real option holder has a flexibility option (call option) on operating the power generation asset. Williams Companies and other power companies such as ExxonMobil, Dynegy, and Noram Energy have executed a number of these long-term tolling arrangements since the late 1990s to provide wholesale power in the United States.

This real option investment can be viewed as a portfolio of short-term spread options on the spark spread (i.e., the difference between electricity prices and natural gas prices). Simply put, the power plant can be operated when the spark spread is in the positive and turned off when the spark spread is in the negative.[8] Consider two scenarios: Scenario 1 uses a positive spark spread (i.e., the option is "in the money") and Scenario 2 illustrates a negative spark spread (i.e., the option is "out of the money"). Both scenarios assume an efficient power plant (higher heat rate, which is the thermal efficiency). The formula for the spark spread is:

$$\text{Spark Spread} = \text{Price of Power per MWh} - [\text{Price of Natural Gas per MMBtu} \times (\text{Heat Rate}/1{,}000)] \quad (36.2)$$

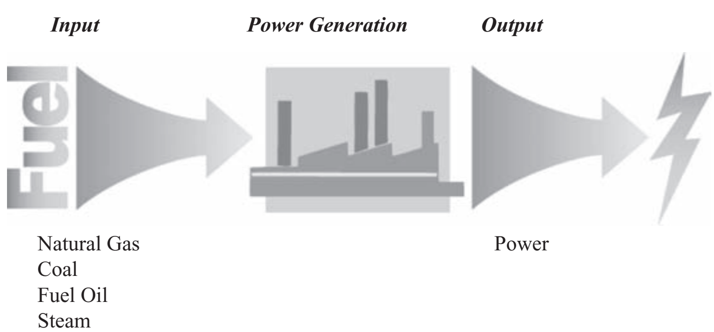


Exhibit 36.3 Simple Illustration of the Tolling Concept

Scenario 1: Positive Spark Spread. For example, assume natural gas prices equal \$5.00/million Btu (MMBtu) and power prices equal \$45.00. Given a heat rate of 8,500 (which is quite efficient), the spark spread equals \$2.50 per megawatt (MWh).

$$\text{SparkSpread} = \$45.00 - [\$5.00 \times (8{,}500/1{,}000)] = \$45.00 - \$42.50 = 2.50$$

The positive spark spread indicates that it is profitable to operate the power plant.

Scenario 2: Negative Spark Spread. Assuming the same natural gas price, a lower power price of \$35.00, and the same heat rate, the spark spread equals a loss of \$7.50 per MWh:

$$\text{SparkSpread} = \$35.00 - [\$5.00 \times (8{,}500/1{,}000)] = \$35.00 - \$42.50 = (\$7.50)$$

The negative spark spread indicates that it is unprofitable to operate the power plant.

Example: Flexibility of Inputs with Dual-Fired Power Plants

As mentioned, production systems that can change easily from one input to another have an embedded real option. Thus, dual-fuel power plants that can use either oil or gas give the operators of these assets the right to switch between fuels whenever it is economical to do so (i.e., the switching option is in the money). For these plants, the owner can be viewed as having a call option on oil and a put option on gas (versus oil only) and the risk increases (decreases) if inputs are positively (negatively) correlated.[9]

Example: Valuing a Natural Gas Storage Facility

Real option analysis can be applied to valuing natural gas storage facilities because for certain types of storage, such as a salt dome caverns, there is substantial flexibility in the operations. Using salt caverns, the natural gas can be more easily withdrawn and sold when prices are high (or to take advantage of short-term price spikes) or injected into storage when prices are low. Also, the facilities can be operated to trade around expected future gas prices and can maximize real option opportunities. Such options can be very valuable based on the seasonality and volatility in natural gas prices. Therefore, owners or leasers of storage can buy in cheap seasons and sell in expensive seasons and concurrently trade on derivative contracts on natural gas. A gas storage facility (again ignoring operating characteristics) can be thought of as a series of call and put options of different strikes. Depending on the amount of gas in storage and the characteristics of the facility, the relative influence of the put and call components vary.[10]

Option to Abandon or Temporarily Shutdown

The right to abandon or scale back an investment has value to the firm when its cash flows do not measure up to expectations. Such options allow the firm to shut down a financially harmful project. As illustrated in Exhibit 36.4, the firm (holder of the option) has a put option on the project's value. If the project is not satisfying

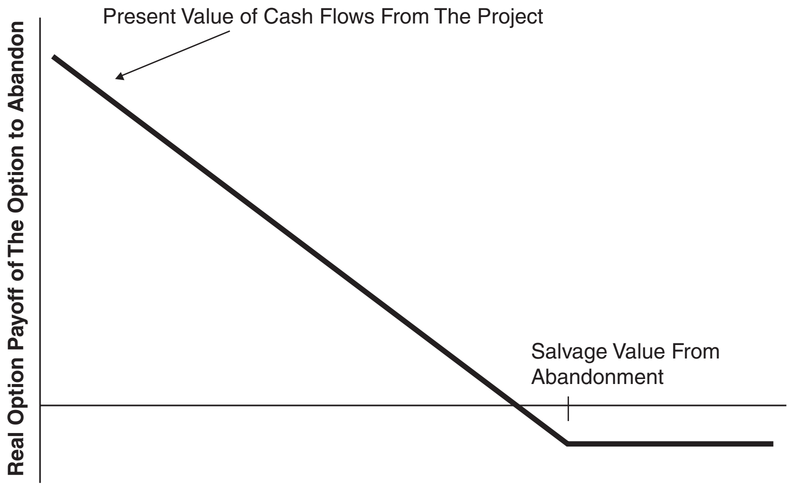


Exhibit 36.4 Option to Abandon a Project

the goal of maximizing shareholder value, the put option can be exercised and the project and its assets divested. With active management throughout the project's life, the choice can be made to shut down in any period when there would be an operating loss (i.e., value of the output is less than the cost to produce).

Example: The Option to Shutdown a Stripper Well

An example of an abandonment option is the decision of whether to shut down a stripper well (an oil well that produces low volumes of oil) that is marginally profitable. Because these wells produce low volumes, frequently less than 1,000 barrels a year, the operating cost per barrel is high and the profitability is very sensitive to oil prices. When oil prices dropped substantially in 1998, many stripper wells in Oklahoma and Texas became unprofitable and were shut down. This decision can be viewed as permanent because shutting down production may cause the geology to change (oil formation collapse or similar), making these decisions irreversible ones.[11]

Example: Option to Temporarily Shutdown a Fertilizer Plant

Natural gas is the primary feedstock (i.e., input) for nitrogen fertilizer (anhydrous ammonia) and comprises approximately 80 percent of the cost.[12] Furthermore, fertilizer is a world market commodity, impacted by supply and demand around the world. Due to high volatility in both natural gas and fertilizer prices, the real option to temporarily shut down a fertilizer plant has significant value. The option to temporarily shut down a production operation can be viewed as a series of put options because the option to shutdown can be exercised at any time up until the expiration date (life of the plant). The option premium is the expense of terminating operations for a period (such as engineering costs to mothball equipment, associated administrative costs, and subsequent start-up costs or other incremental expenses). The strike price is the difference between fertilizer prices and the cost of operating (approximately 80 percent of which is due to the price of natural gas).[13] This option is most valuable when there is an extended period of high natural gas prices such as has occurred over the period 2003 through 2008.

Hybrid Real Options

More recently, researchers have combined real options theory with other theories, hence the name *hybrid real options*, in order to perform a more comprehensive and realistic analysis of complex problems. Dias (2005) applies this technique specifically to the petroleum industry. The two main combinations under hybrid real options are:

1. The combination of real options theory with game theory to consider endogenously the strategic behavior of other firms.
2. The combination of real options theory with methods from probability theory and Bayesian statistical decisions.

The models under this technique generate a new way to model technical uncertainty of a project in dynamic real options models. In essence, using hybrid real options, these two combinations can be recombined so as to capture the value of information differences in noncooperative and cooperative games.

For example, in the petroleum industry when using hybrid options, important variables like exploratory chance factor, volume, and quality of a petroleum reserve, can be modeled with the development of a new theory on revelation distribution and measures of learning.[14] Dias (2005) describes the combination of real options theory with the evolutionary computation theory (evolutionary real options) and states they hold great potential in complex applications of optimization under uncertainty. He presents this method to illustrate an application using the genetic algorithms to evolve the decision rule for optimal exercise of a real option.

VALUING REAL OPTIONS

Applying real option valuation to energy assets involves these challenges:

- Modeling of energy prices.
- Modeling of energy demand.
- Understanding the relationships between forward prices and forecast prices.
- Understanding liquidity in derivatives markets.

As discussed earlier, the value of a project with real option features is equal to the project's traditional NPV plus the real option value. How can a real option value be quantified? Just as in NPV analysis, real option valuation involves making a number of assumptions based on uncertainty. But it is this very uncertainty that gives the real option value and the greater the uncertainty, the greater the potential value. It is important to calculate a specific value as a starting point, but remember that the computed value is just an estimate. Next, methods to value real options are described. There is an extensive literature on valuing real options.[15]

Decision Trees

The use of a decision tree seems most in line with the needs of managers, although more sophisticated methods are available. This approach is, in fact, the simplest approach and can be used to examine projects which have embedded options to abandon, expand, contract, defer, or extend.

The basic decision tree can be explained in this way: The project under consideration will have a net investment at time zero. Following this net investment, or cash outflow, the decision maker will assign probabilities to different possible future states of nature. For example, there may be three possible futures states: low demand, medium demand, and high demand. Each future state of nature has a corresponding expected cash inflow. Obviously, cash inflows will be higher when demand is highest and cash inflows will be lowest when demand is lowest. A finite set of possible states is used for simplicity and different NPVs are calculated based on the cash flows and the probability of those cash flows, along with an appropriate discount rate. The decision maker is then left with multiple NPVs and their corresponding probabilities. This is an example of a decision tree without any embedded options.

The value of real options is seen when embedded options are introduced into the decision tree. Introducing an option to defer will allow the decision maker to make the initial investment at a time other than today. This will change the cash flows and also can reduce uncertainty. Similarly, options to expand or abort will also impact cash flows and reduce risk while also increasing the expected NPV. The ability of embedded options to increase the expected NPV, while at the same time reducing uncertainty, creates value.

Monte Carlo Simulation

It should be noted that while decision tree analysis is prevalent in capital budgeting decisions that incorporate real options, there is an obvious shortcoming. Decision tree analysis offers only a finite amount of possibilities for future states. To incorporate multiple risks and an almost infinite amount of possible future states, a Monte Carlo simulation can be used. The decision maker is then provided with a single measure of volatility for valuing real options. The result of the simulation is a probability distribution in which different outcomes and their likelihoods of occurrence are presented. In this manner, an NPV is calculated that has incorporated many different courses of action in the future—that incorporates the use of real options.

Option Pricing Models

Finally, real options can be valued using the Black-Scholes (B-S) pricing model. Five inputs are needed to price an option (including real options) using B-S:

1. Current price
2. Exercise price
3. Time to maturity
4. Risk-free rate
5. Volatility

Given these inputs, the value (or price) of a real option can be calculated. The difficulty, as with pricing any option, is in quantifying the volatility. Furthermore, modeling energy prices is extremely difficult and entire books have been written on this topic.[16]

CONCLUSION

The application of real option analysis is more than just another capital budgeting tool; it is a way of thinking. Traditional net present value analysis does not capture the value of real options embedded in projects. The types of real options discussed in the chapter include the option to expand, the option to wait, the option to vary production inputs, outputs, or processes, the option to abandon or temporarily shutdown, and hybrid real options. These real options were illustrated using applications in the energy industry. As noted earlier, the value of these real options is dependent on management insight and strategic decision making. If properly valued and exercised, real options add substantial value to projects. Likewise, if these flexibilities are not properly exercised or worse yet, if they are ignored, the value of real options embedded in projects will not be fully realized.

Real options are still not widely used in practice. Applications are found most commonly in the pharmaceutical and oil and gas exploration industries—two industries characterized by high risk. Two possible reasons that real option analysis is still not widely adopted are the complexity of real option analysis and the lack of corporate incentives to undertake projects justified by this analysis. Clearly, there is still a lack of understanding about real option applications in many industries.

More research is needed on the application of real options. There is still a lack of good empirical work on real options. As Copeland, Weston, and Shastri (2005) note:

> *Even though there has been a great deal of successful research into the pricing of all types of financial options, there has been much less work on real options. . . . We note that no one has provided any evidence that real options are priced correctly in the marketplace. Both empirical evidence and experimental evidence are needed. (p. 871–872)*

ENDNOTES

1. Crystal Ball is produced by Oracle and is a leading spreadsheet-based application for predictive modeling, forecasting, simulation, and optimization.
2. There is a robust literature on applying real options to capital budgeting decisions. Alex Triantis has a searchable database of over 500 practitioner and academic articles on real options. Refer to www.smith.umd.edu/faculty/atriantis/ for more information. Also refer to Dixit and Pindyck (1994), Damodaran (2000 and 2008), and references listed in Exhibit 36.1 of this chapter.
3. The authors have interviewed a number of business professionals and were surprised to how many were unfamiliar with the term *real options*.
4. Dixit and Pindyck (1994, chap. 1) present a comparison between the real options approach and the neoclassical investment theory with their variants, such as the Tobin's q or marginal q and the Jorgenson's user cost of capital. Both variants of the neoclassical model rely on the well-known net present value rule.

5. There are three distinct phases of oil recovery: primary, secondary, and tertiary (or enhanced) recovery. Primary recovery can be viewed as the easiest because the natural pressure of the reservoir or gravity drive oil into the wellbore and then artificial lift techniques, such as pumps, can bring the oil to the surface. Some estimates indicate that only about 10 percent of a reservoir's original oil in place is typically produced using primary recovery. Secondary recovery techniques typically involve injecting water or gas to displace oil and result in the recovery of 20 to 40 percent of the original oil in place. After using primary and secondary techniques, some producers utilize several tertiary, or enhanced oil recovery (EOR), techniques, which commonly involve thermal recovery, gas injection, and chemical injection. These techniques can recover 30 to 60 percent, or more, of the reservoir's original oil in place. However, these techniques are much more expensive and depend on the petroleum economics.
6. For more information, see McCormack and Sick (2001), Ronn (2005b), among others.
7. Note: The spark spread represents the theoretical margin for a power plant. If the price of electricity (the output) is higher than the price of fuel (such as natural gas) for a power plant, then the spark spread is positive. Likewise, if the price of electricity is less than the cost of the fuel, the spread is negative.
8. A number of complex factors need to be incorporated into the modeling of this type of real option. Plant physical constraints in the form of start-up costs, ramp-up and ramp-down costs, and maintenance due to wear and tear associated with switching the plant on and off, must be incorporated into the real option valuation analysis. Furthermore, these facilities can run at different capacity usage amounts. A number of articles have been written on this topic. For example, see Leppart (2005), Pilipovic and Wengler (1998), and Price (1997), among others.
9. For more discussion, refer to Kulatilaka (1993) and Ronn (2005a).
10. For more information on valuing real options on natural gas storage, refer to De Jong and Walet (2005) and Thompson, Davison, and Rasmussen (2007), among others.
11. Refer to Titman and Martin (2008, pp. 466–468) for additional discussion on this topic.
12. It takes 34,000 million Btus of natural gas to make 1 ton of anhydrous ammonia.
13. It is important to note that the volatility of natural gas historically has been significantly higher than the volatility of most other commodities, including crude oil and fertilizer. Additionally, natural gas prices have a strong seasonal component.
14. As Dias notes: "The real options theory is combined with other theories—so the name hybrid real options—in order to perform a more comprehensive and realistic analysis of complex problems that arises from petroleum industry. The two main combinations analyzed here are: (a) the combination of real options theory with game theory—real options games—to consider endogenously the strategic behavior of other firms, especially in the optimal stopping game with positive externalities known as war of attrition, as well as the possibility to change this game by a cooperative bargain game; and (b) the combination of real options theory with methods from probability theory and Bayesian statistical decision—Bayesian real options—generating a new way to model technical uncertainty of a project in dynamic real options models. These two combinations are re-combined in order to obtain an adequate solution that captures the value of information differences in non-cooperative and cooperative games." See www.puc-rio.br/marco.ind/abstract.html for the quote in English. The dissertation is in Portuguese.
15. For more information on valuing real options, particularly in the energy industry, refer to references listed in Exhibit 36.1 and also to Amran and Kulatilaka (1999), Kaminski (2005), Ronn (2005a), and Titman and Martin (2008).
16. Risk Books (London) has published a number of excellent books on modeling energy prices and related issues. Refer to http://riskbooks.com/ for more information.

REFERENCES

Amram, M., and N. Kulatilaka. 1999. *Real Options: Managing Strategic Investment in an Uncertain World*. Boston: Harvard Business School Press.

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81: 637–654.

Brennan, M. J., and E. S. Schwartz. 1985. "Evaluating Natural Resource Investments," *Journal of Business* 58, No. 2: 135–157.

Copeland, T. E., and V. Antikarov. 2001. *Real Options: A Practitioner's Guide*. New York: Texere.

Copeland, T. E., J. F. Weston, and K. Shastri. 2005. *Financial Theory and Corporate Policy*, 4th ed. Boston: Pearson Education.

Damodaran, A. 2000. "The Promise of Real Options," *Journal of Applied Corporate Finance* 13, No. 2: 28–39.

Damodaran, A. 2008. "The Promise and Peril of Real Options," Working paper, New York University; available at www.ssrn.com.

Dawson, T., and J. Considine. 2003. "Real Option Valuation and Equity Markets," *Risk* (May): 98–100.

De Jong, C., and K. Walet. 2005. "Gas Storage Management," in V. Kaminski, ed., *Managing Energy Price Risk: The New Challenges and Solutions*, p. 631–648. London: Risk Books.

Dias, M. A. G. 2005. *Hybrid Real Options with Applications in Petroleum*, PhD diss., Industrial Engineering Department, PUC-Rio, Rio de Janeiro, Brazil.

Dixit, A. K., and R. S. Pindyck. 1994. *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

Grenadier, S., ed. 2000. *Game Choices: The Intersection of Real Options and Game Theory*. London: Risk Books.

Kaminski, V., ed. 2005. *Managing Energy Price Risk: The New Challenges and Solutions*, London: Risk Books.

Kasanen, E., and L. Trigeorgis. 1993. "Flexibility, Synergy, and Control," in R. Aggarwal, ed., *Strategic Investment Planning, Capital Budgeting under Uncertainty*, p. 208–231. Englewood Cliffs, NJ: Prentice-Hall.

Kulatilaka, N. 1993. "The Value of Flexibility: The Case of Dual-Fuel Industrial Steam Boiler," *Financial Management* 22, No. 3: 271–280.

Leppart, S. 2005. "Valuation and Risk Management of Physical Assets," in V. Kaminski, ed., *Managing Energy Price Risk: The New Challenges and Solutions*, p. 649–682. London: Risk Books.

Martin, A., D. A. Rogers, and B. J. Simkins. 2004. "Southwest Airlines: The Blended Winglet Project," *Case Research Journal* 24 (Summer/Fall): 23–33. (Published by the North American Case Research Association.)

McCormack, J. M., and G. Sick. 2001. "Valuing PUD Reserves: A Practical Application of Real Option Techniques," *Journal of Applied Corporate Finance* 13, No. 4 (Winter): 110–115.

Miller, K., and H.,G. Waller. 2003. "Scenario, Real Options and Integrated Risk Management," *Long Range Planning* 36: 93–107.

Myers, S. 1977. "Determinants of Capital Borrowing," *Journal of Financial Economics* 5: 147–175.

Pilipovic, D., and J. Wengler. 1999. "Basis for Boptions," *Energy & Power Risk Management* (December/January): 28–29.

Price, J. P. 1997. "What's a Power Plant Worth?" *Public Utilities Fortnightly*, September 15, 38–41.

Ronn, E. I. 2005a. "Valuation of Oil Fields as Optimal Exercise of the Extraction Option," in V. Kaminski, ed., *Managing Energy Price Risk: The New Challenges and Solutions*, p. 609–630. London: Risk Books.

Ronn, E. I., ed. 2005b. *Real Options and Energy Management: Using Options Methodology to Enhance Capital Budgeting Decisions*. London: Risk Books.

Smit, H., and L. Ankum. 1993. "A Real Options and Game-Theoretic Approach to Corporate Investment Strategy under Competition," *Financial Management* (Autumn): 241–250.

Thompson, M., M. Davison, and H. Rasmussen. 2007. "Natural Gas Storage Valuation and Optimization: A Real Options Application," Working paper, University of Western Ontario.

Titman, S., and J. D. Martin. 2008. *Valuation: The Art & Science of Corporate Investment Decisions*. Boston: Pearson Education.

Triantis, A. J. 2000. "Real Options and Corporate Risk Management," *Journal of Applied Corporate Finance*, 13, No. 2 (Summer): 64–73.

Triantis, A. J., and A. Borison. 2001. "Real Options: State of the Practice," *Journal of Applied Corporate Finance* 14, No. 2 (Summer): 8–24.

Trigeorgis, L. 1993a. "The Nature of Option Interactions and the Valuation of Investments with Multiple Real Options," *Journal of Financial and Quantitative Analysis*. 28, No. 1: 1–20.

Trigeorgis, L. 1993b. "Real Options and Interactions with Financial Flexibility," *Financial Management*. 22, No. 3: 202–224.

ABOUT THE AUTHORS

Betty J. Simkins is Williams Companies Professor of Business and associate professor of finance at Oklahoma State University. She received her BS in chemical engineering from the University of Arkansas, her MBA from OSU, and her PhD from Case Western Reserve University. Professor Simkins is very active in the finance profession and currently serves as president of the Eastern Finance Association, sits on the board of directors for the Financial Management Association (FMA), is coeditor of the *Journal of Applied Finance*, and is executive editor of *FMA Online*. She has coauthored more than 30 journal articles in publications including the *Journal of Finance, Financial Management, Financial Review, Journal of International Business Studies, Journal of Futures Markets, Journal of Applied Corporate Finance,* and *Journal of Financial Research* and has won a number of best-paper awards at academic conferences.

Kris Kemper is a PhD candidate at Oklahoma State University. He received his both his BBA in finance and his MBA from the University of Central Oklahoma. Prior to attending Oklahoma State University, he worked for five years in the financial services industry and also spent four years in the United States Air Force. His current research interests include mutual funds, credit ratings, and capital structure.

CHAPTER 37

Using Derivatives to Manage Interest Rate Risk

STEVEN L. BYERS
U.S. Securities Exchange Commission

INTRODUCTION

Interest rates are volatile, exhibiting a range of movement that can be only partially explained by various historical and economic factors, leaving a great deal of uncertainty for financial and nonfinancial entities with large risks that need to be managed. Corporations with debt face volatile and uncertain payment schedules, banks with lending portfolios have to be able to fund their fixed rate activities by hedging floating rate products, funds and others with bond portfolios must manage their risk to ensure that they can meet current and future payout obligations all the while meeting performance benchmarks for the portfolio. Because of these activities, there is a huge demand for products that can be used to manage risks associated with interest rate fluctuations.

Basically two building block derivative products exist to manage these risks: forward-based instruments and options. The plethora of interest rate derivatives available are primarily variations and modifications of these building blocks. Instruments in the forwards classification include forward rate agreements, interest rate futures contracts, and interest rate swaps. The options classification consists of option contracts on forward-based instruments, futures options, as well as options on spot interest rate instruments including caps and floors.

FORWARD-BASED INSTRUMENTS

Forward Rate Agreements

The concept of the forward rate agreement (FRA) is crucial to understanding interest rate derivatives. Therefore, we will explain in detail the mechanics in order to build the basis for futures and options. A forward rate agreement is an agreement between two parties; one party, the FRA buyer, is a notional borrower, whereas

The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The view expressed herein are those of the author and do not necessarily reflect the views of the Commission or of the author's colleagues upon the staff of the Commission.

the FRA seller is a notional lender. The notional amount (or notional principal amount or notional value) on a financial instrument is the nominal or face amount that is used to calculate cash flows made on that instrument. The notional loan covers a period in the future. The details of the notional loan include: the amount of principal, currency, fixed interest rate, and the start and maturity dates. Why use FRAs? Counterparties may wish to obtain protection from future interest rate movements, a borrower may fear rising rates, and lender may fear falling rates. Or they may wish to speculate on future interest rate movements. FRAs are used more frequently by banks, for applications such as hedging their interest rate exposures, which arise from mismatches in their interest-sensitive portfolios. FRA's are also used widely for speculative activities. They are advantageous because they do not involve credit risk on the principal amount, they do not tie up credit lines for extended periods, they are off-balance sheet instruments (advantageous for entities—banks, for instance—subject to regulatory capital requirements in that the capital requirements pertain to the balance sheet).

The two key features of the FRA include:

1. The FRA fixes the interest rate for both parties: The buyer fixes the interest rate for borrowing or funding, the lender fixes the interest rate for lending or investing.
2. The FRA is a notional loan, meaning that no actual lending or borrowing takes place. Rather the protection is provided by a settlement sum that is eventually paid by one party to the other to compensate for the difference between the fixed rate originally agreed to under the FRA and the actual rate prevailing in the market.

Let us now look at a simple example to illustrate the basic mechanisms of the FRA. Caribou Research, Inc. (the company) needs to borrow a principal amount of $1 million for 90 days starting in three months. To protect against interest rate risk, the company considers an FRA covering the six-month period starting three months from now. This is referred to in the market as a "3-against-6-month" FRA or in simpler terms a 3×6 FRA. Let us assume the market rates are 4.75 percent for three months, and banks are quoting a rate of 5.00 percent for a 3×6 FRA. The company buys a 3×6 FRA on $1 million at a contract rate of 5.00 percent. In three months, let us assume interest rates have risen to 6.00 percent. The bad news is that the company will have to borrow at the higher prevailing rate, which results in payment of an extra $2,500 interest paid. The good news is that the FRA settlement sum will provide the company with compensation for the extra interest paid. The settlement sum is calculated using this formula:

$$\textit{Settlement Sum} = \frac{A \times (i_r - i_c) \times \frac{L}{B}}{1 + \left(i_r \times \frac{L}{B}\right)} = \$1{,}000{,}000 \times \frac{(.06 - .05)}{\left[\frac{360}{90} + .06\right]} = \$2{,}463$$

where $A = \$\,1$ million, the contract amount
$i_r = 6\%$, the reference rate
$i_c = 5\%$, the contract rate
$L = 90$, the length of the contract period in days
$B = 360$, days per year assumed in interest calculations (in this case the US dollar convention works on a 360-day year)

Clearly Caribou Research (buyer of the FRA) was protected by the FRA settlement sum of $2,463 against the additional interest it owed, $2,500 when rates increased to 6 percent. The seller of the FRA is the losing party in this scenario. But what would have happened if rates had gone down? For the same FRA contract but with the reference rate, i_r, decreasing in three months' time to 4 percent, a settlement amount of $2,475 would be paid by the borrower to the seller. The general case for FRAs when market interest rates change is:

- The buyer of the FRA is protected against rising rates (but pays if rates fall).
- The seller of the FRA is protected against falling rates (but pays if rates rise).

Hedging with FRAs

Suppose on Friday, June 1, 2001, Telluride Bank has negotiated a three-month $1 million loan to a large corporate client, Caribou Fine Woodworking, at a fixed rate of 5.75 percent. The loan is to be drawn down on Wednesday, September 5, 2001, with a rate fixed on September 3, and is to be repaid on Wednesday, December 5, 2001. The bank decides to protect its margin by buying a 3 × 6 FRA, currently quoted at 5.50 percent, in order to secure a 0.25 percent profit margin. Then assume that on September 3, 2001, the three-month LIBOR turns out to be 6.5 percent. The settlement sum is calculated to be $2,486.92, which the bank then invests at 6.5 percent for 91 days to yield $2,527.78. In doing so, the bank has locked in a profit margin of $631.94. The profit margin is the equivalent to a 0.25 percent margin on $1 million for 91 days, showing that the FRA provided a perfect hedge. A summary of the transaction is:

Interest receivable on loan @ 5.75%	$14534.72
Interest payable on financing @ 6.5%	–$16430.56
Loss on loan due to rise in LIBOR	–$1895.83
Proceeds from FRA@ 5.5% at maturity	$2527.78
Total profit resulting from gain on hedge	$631.94

Interest Rate Futures Contracts

Interest rate futures were developed in the aftermath of the collapse of the fixed exchange rate regime in the early 1970s, which led to increased volatility in interest rates. The Chicago Board of Trade (CBOT) launched the first interest rate futures contract in 1975. Since then, the menu of interest rate futures contracts has expanded to include many types of contracts written on a wide variety of interest rates. One of the most popular contracts is the Eurodollar futures contract.

A Eurodollar futures contract is essentially a standardized, exchange-traded FRA. It is a standardized commitment that is traded primarily on the Chicago Mercantile Exchange (CME) and the Euronext. The term *standardized* means that certain terms and conditions for the futures contracts must be established in order for them to be easily traded and cleared. The main terms that must be standardized are:

- Underlying reference interest rate
- Contract size (nominal face value)
- Price quote convention

- Minimum price movement or tick size
- Contract delivery months
- Asset delivery mechanism
- Last trading day for the contract

The market for a futures contract is highly liquid, which makes it easy to establish and liquidate positions. The whole process is transparent, and the current menu of products allows managers to manage interest rate risk for as little as 1 day or as long as 10 years. With the FRA, the risk protection is effective only if the counterparty to the agreement does not default; futures exchanges, however, mitigate the counterparty credit risk by having their trading operations linked to a financially sound clearinghouse.

The clearinghouse acts as a buyer to every seller and a seller to every buyer; thus, the credit risk for each transaction lies with the clearinghouse. In addition to the clearinghouse's interdisposition between the parties who execute transactions through the exchange, all trading is done against a margin account, which acts as a performance bond. A party to a futures contract must post a prescribed amount of margin before they can enter into a futures trade. All margin accounts are marked to market at the end of every trading day. If the market moved favorably (unfavorably) for one party, its margin account is credited (debited) with the amount by which the market price moved in its favor. If the market move is such that there are insufficient funds in the margin account, the debited party must restore the margin level by adding additional funds to the margin account prior to the next trading day. The arrangement between the trading exchange and the clearinghouse drastically reduces the counterparty credit risk of the futures contract.

Quotation and Pricing Convention for an Interest Rate Futures Contract

Interest rate futures are traded on an indexed "price" rather than the interest rate itself, where the price is defined as

$$P = 100 - i$$

where P = price index
i = future interest rate in annualized percent

This method of quotation reflects the inverse relationship between the price and yield, and it allows traders to follow the conventional wisdom of buying low and selling high. For example: 6.37 percent equals an index price of 93.63; should rates fall to 5.5 percent, the index price will rise to 94.5. The futures price is not a monetary measure, meaning that a futures price of 94.50 does not imply that the price is $94.50. The futures price is an alternative representation for the interest rate at which an underlying notional deposit or loan could be executed. It is a proxy for the general level of interest rates in the same manner as a stock index such as the Standard & Poor's 500 index is just an indicator of the level of the stocks that make up the index.

Basis Risk

Interest rate futures allow a user to lock into a rate for future transaction. However, it is not today's cash market rate that they are able to lock in but rather the forward rate. The difference between today's cash market rate (spot price) and the forward rate at some particular date in the future is called the price basis. As the last trading date of the futures contract approaches, cash prices and futures prices move closer together. This narrowing of the gap between the two prices continues to converge until it is theoretically reduced to zero, the time when the futures month is the spot month. This is the case because the futures price will no longer reflect a forward rate; instead, it is the rate for a deposit made now. Arbitrage between the cash rate and the futures price ensures that they must be equal at expiration.

The price basis is described as being positive or negative. It is customary to calculate it as the spot price less the futures price. Therefore, a basis is positive when the spot price is greater than the futures price, and it is negative when the spot price is less than the futures price. During the life of the hedge, it will be influenced positively or negatively, but what is most important is that the basis will narrow continuously until expiry of the futures contract month.

A second type of basis is called the quantitative basis. It arises due to differences in the quantity (notional amount) of the underlying traded in the spot market and the quantity (notional amount) of the underlying hedged in the futures market. Recall that futures contracts are standardized into whole-lot sizes that cannot be divided up into smaller amounts to exactly match size of the exposure to be hedged. Quantitative basis does not change; it is locked in when the futures contract is entered into. The result is that hedge is either overhedged or underhedged. By under- (over-) hedging, the user is using fewer (more) futures contracts to hedge than is necessary to equal their quantitative exposure in the cash market.

Futures Hedge Ratio

Because a futures contract is standardized, the terms of the contract may not match the terms of the nominal lending or borrowing of the entity wishing to hedge with futures. Since the purpose of hedging is to reduce the interest rate risk exposure, it is necessary to choose the number of futures contracts used. A perfect risk-minimizing hedge would be one where every dollar nominally borrowed or lent is exposed to the opposite risk in the futures market. Such a hedge would produce a hedge ratio of 1:1. However, this may not always be possible because of the standardized notional contract sizes for interest rate futures. It is therefore necessary to calculate the risk-minimizing hedge ratio h, which will lead us to the calculation for the optimal number of futures contracts C.

The risk-minimizing hedge ratio h, is calculated as:

$$h = \rho \frac{\sigma_S}{\sigma_F}$$

where $\rho =$ coefficient of correlation between the standard deviation of the change in the spot price, S, and the change in the futures price, F, during a period of time equal to the life of the hedge

$\sigma_S =$ standard deviation of the change in the spot price, S

$\sigma_F =$ standard deviation of the change in the futures price, F

The risk-minimizing hedge ratio, h, is equivalent to the slope of the best fit line when the change in the spot price, S, is regressed on the change in the future price, F, over a period of time equal to the life of the hedge.

The optimal number of futures contracts, C, required for the risk-minimizing hedge is then calculated by:

$$C = \frac{hN_a}{Q_f}$$

where $h =$ risk-minimizing hedge ratio
$N_a =$ size of the position being hedged
$Q_f =$ size of one futures contract

Example of Hedging with Eurodollar Futures

In April, Caribou Fine Woodworking is being sold for \$4,250,000, and the contract will be finalized in December. The owners intend to purchase another company sometime after the sale is finalized so they want to deposit the funds for 12 months and receive quarterly interest payments. At this time, the current spot rate is 4.76 percent, the trend in spot interest rates is declining, and it is predicted to continue so the owners decide to lock in the rate they will receive. The owner finds that December Eurodollar futures on the Chicago Mercantile Exchange are trading at 95.16, which implies an annualized interest rate of 4.84 percent. The owner calculates that the risk-minimizing hedge ratio, h, is 0.92. The face value of each futures contract is \$1,000,000 so the optimal number of futures contracts, C, is:

$$C = \frac{hN_a}{Q_f} = 0.92\,\frac{\$4{,}250{,}000}{\$1{,}000{,}000} = 3.91$$

The owners enter four futures contracts, since they are available only in integer amounts, to hedge the \$4,250,000 exposure in the cash market, leaving them with an overhedged position and a negative basis.

In November, just when the owners are about receive the proceeds from the sale of the company and are ready to deposit the money, just as was predicted, interest rates have continued to decline and the three-month London Interbank Offered Rate (LIBOR) is now 4.015 percent (annualized). The owners lift the hedge by selling the futures contracts at 95.98, with an implied interest rate of 4.020 percent (annualized). The trade in the futures market results in a profitable hedge. We see this by analyzing the gain in the futures trade and the loss in the anticipated interest income in the cash market. The futures trade results in a gain of \$8,200 per contract, or \$32,800 in total, because the sale of the futures contracts took place at a higher price, 95.98, than the purchase price back in April of 95.16, thus gaining 0.82 contract points, or 0.820 percent, during the term of the hedge. While the owners' anticipated interest income from depositing the money and receiving quarterly interest payments declined from \$50,575 (4.76% $\times$ \$4,250,000/4) to \$42,659.37 (4.015% $\times$ \$4,250,000/4) for a quarterly "loss" of \$7915.625, or \$31,662.50 in total for the year. The \$32,800 profit on the futures trade less the loss on the anticipated interest income of \$31,662.50 results in a net profit of \$1,137.50. This net profit is the result of the negative quantitative basis achieved by overhedging.

Hedging a Portfolio of Coupon Bonds with Interest Rate Futures

Hedging a portfolio of coupon bonds requires that coupon and maturity mismatches between assets and liabilities on the balance sheet be taken into consideration (e.g., a bank manager needs to take into account the maturity of the bonds [assets] and any planned withdrawals or distributions [liabilities] by depositors). A common approach to manage the mismatch is to employ a duration-based hedging strategy.

The duration of a financial asset measures the sensitivity of the asset's price to interest rate movements, expressed in years. The reason for expressing this sensitivity in years is that the time that will elapse until a cash flow is received allows more interest to accumulate. Therefore, the price of an asset with long-term cash flows has more interest rate sensitivity than an asset with cash flows in the near future. This relationship leads to one common measure of duration that is calculated as the weighted average number of years to receive each cash flow. This can written as:

$$Duration = \sum_{i=1}^{n} t_i \left[\frac{c_i e^{-y t_i}}{B} \right]$$

where $c_i =$ the cash flow from the bond at time
$t_i =$ the timing of the cash flows
$B =$ the bond price
$y =$ the yield (continuously compounded)

For example, a zero-coupon bond maturing in five years has a duration of five years; however, a coupon bond maturing in five years has a duration of less than five years due to the receipt of cash payments (coupons) prior to five years.

Duration applies only to small changes in interest rates, as it is a linear measure of how the price of a bond changes in response to interest rate changes. However, as interest rates change, the price does not change linearly but rather is a nonlinear, convex function of interest rates. This factor is known in finance as convexity, and it is a measure of the curvature of how the price of a bond changes as the interest rate changes for moderate or large changes in interest rates. Convexity is calculated as:

$$Convexity = \frac{\sum_{i=1}^{n} c_i t_i^2 e^{-y t_i}}{B}$$

The convexity of a bond portfolio typically is greatest when the cash flows from the portfolio are spread evenly over a long period. Likewise, convexity tends to be smallest when the payments are concentrated around one particular point in time. When convexity is hedged the portfolio is protected against large parallel shifts in interest rates.

Duration and convexity make up a large topic; there are numerous measures of duration involving varying levels of complexity. For a comprehensive discussion of duration, see Bierwag (1987).

The duration of a bond portfolio is defined as the weighted average of the durations of the individual bonds held in the portfolio, with the weights being proportional to the bond prices. A portfolio of coupon bonds can be hedged effectively using interest rate futures; at issue is calculating the correct number of interest rate futures contracts. Using duration, the risk manager can determine the number of interest rate futures contracts required to hedge against movements in interest rates. This is referred to as the duration-based hedge ratio; it is sometimes called the price sensitivity hedge ratio. An excellent example of this approach is given in Kolb and Overdahl (2007).

Interest Rate Swaps

A swap is an agreement between two parties to exchange cash flows on particular dates in the future. A common interest rate swap is the fixed-for-floating arrangement where one counterparty to the contract agrees to pay fixed rate interest payments in exchange for floating rate interest payments on a predetermined set of future dates. The interest payments are based on a notional principal and are calculated on different bases. One party is called the fixed rate payer (or floating rate receiver), and the other party is the floating rate payer (or fixed rate receiver). At initiation, the present value of the fixed rate payments will be equal to the expected value of the floating rate payment, so the swap has no value (trades at par).

Hedging a Term Loan with an Interest Rate Swap

Caribou Research Inc. is preparing to embark on a lengthy research project requiring it to secure a loan for $24,500,000 in three months. Telluride Bank quotes the company a fixed rate loan for four years at 6.1 percent per annum or a floating rate loan of four years at a rate of 0.80 above six-month USD LIBOR with LIBOR currently at 5.2 percent. Caribou must make interest rate payments every six months with the first payment coming due after the first six-month period.

Caribou opts for the floating rate loan. Because of this, Caribou faces the risk that LIBOR rates will rise over the next four years, thus increasing the cost of financing the research project. After one month, Caribou's chief financial officer realizes that the current spot LIBOR yield curve is rising steeply, indicating that the market expects there to be a substantial rise in interests during the life of the loan. She therefore determines that it is in the company's best interest to go to the money markets to structure a swap that will fix its interest rate risk exposure.

Caribou wishes to enter a swap whereby it receives six-month LIBOR and pays fixed interest of 5.24 percent on the notional value of the swap. It agrees on a swap with an investment bank that begins and ends on the same day, and for an equivalent notional as Caribou's bank loan. In this case, the size of the swap and the loan match perfectly and all the payment periods coincide (which is often not the situation but we make this assumption for simplification).

Although Caribou has now locked the rate at which it will make semiannual interest payments, it faces the risk that the six-month LIBOR might decline during

Exhibit 37.1 Liability Linked Float to Fixed Swap

the period of the swap, which is exactly the opposite risk it faces with the loan at the bank. Because the bank loan and the swap coincide exactly, the risks cancel each other out exactly, and Caribou is immunized against interest rate risk.

The cash flows among Caribou, Telluride Bank, and the investment bank are shown in Exhibit 37.1. The swap effectively converts the floating rate obligation of LIBOR + 0.80% to a fixed rate obligation of 6.04% (5.24% + 0.80%), which is 0.06 percent less than the original fixed rate loan quoted by Telluride Bank.

Using Interest Rate Swaps to Manage the Duration of a Bond Portfolio

Interest rate swaps also can be used effectively to alter the duration of a bond portfolio without physical reallocation of assets. Portfolio managers typically shorten the overall duration during times of increased volatility or other market uncertainty, as this will limit the scale of potential losses. The shorter the duration of a bond (or portfolio of bonds), the less the net present value of that bond (or portfolio) will change per basis point change in yield. Unfortunately, during periods of high volatility, rising yields also tend to mean reduced liquidity, making asset reallocation a potentially expensive business, as the portfolio manager is forced to cross wide bid-offer spreads to switch holdings. Interest rate swaps present a solution to this problem, where positions can be established quickly and simply and where the bid-offer spread stays relatively tight even in volatile markets. By entering into an interest rate swap to pay fixed for, say, 10 years against receiving three-month LIBOR, an investor can immediately shorten the duration of his bond portfolio. From a risk perspective, the effect of entering into such a deal is the same as would be achieved by selling 10-year bonds (paying fixed for 10 years) and buying and rolling three-month paper with the proceeds (receiving fixed for three months). Lengthening the duration of a bond portfolio can be achieved in reverse fashion using interest rate swaps. An excellent numerical example of using interest rate swaps to manage the duration of a portfolio containing coupon bonds can be found in Kolb and Overdahl (2007).

OPTION-BASED INSTRUMENTS

Option-based products can be a very attractive tool for managing interest rate risk. Interest rate derivatives have payoffs that are dependent in various ways on the level of interest rates. These instruments are traded on two platforms, over the counter and exchange traded. There is a never-ending and always increasing array of interest rate derivatives on offer to the market; however, the fundamental principles of options are universally the same to all options, no matter where they are traded, what they are called, or what the underlying asset may be. We should point out, however, that interest rate options are much more difficult to engineer than equity or foreign exchange derivatives for a number of reasons, some of which include the complicated behavior of interest rates and the entire term structure of interest rates (the yield curve) and the varying volatility for different maturities on the yield curve, among others. Detailed discussion of these issues is beyond the scope of this chapter but is available in Hull (2008) or in other of the suggested readings.

An interest rate option is an instrument that limits the risk to the buyer of the option. It allows the buyer to profit from a favorable move in the underlying interest rate while giving protection against an adverse move in the underlying interest rate.

As with equity and foreign exchange options, there are two choices of option contract: a call option and a put option. The call option on an interest rate (referred to as a cap) grants the buyer the right to secure the underlying interest rate at a specified price at a specific time (a European option) or within an established time frame (an American option). Though buyers have the right, they do not have the obligation to purchase the underlying interest rate. A put option is many ways the reverse of a call option. The put option on an interest rate (referred to as a floor) grants the buyer the right, but not the obligation, to sell the underlying interest rate at a specified price at a specific time or within an established time frame. The seller of either type of option has the obligation to deliver (for call option) or purchase (for a put option) the underlying rate per the terms of the option contract. The interest rate options we consider are interest rate guarantees (IRGs), caps and floors, collars, and swaptions.

Interest Rate Guarantees

An IRG is an option on a forward rate agreement (FRA) or an interest rate futures settlement rate that provides protection against adverse movements in interest rates while maintaining the upside benefits from favorable interest rate movements. For instance, a company wishes to secure the interest rate it will pay in association for a loan it requires in three months' time. It locks in a rate of 6.4 percent (annualized) at the cost of a premium for the IRG call option. In three months' time, the quoted LIBOR is 6.9 percent. The company thus has saved itself 50 basis points of interest rate expense by having locked in the rate at 6.4 percent. This illustrates the protection feature of the IRG. However, if in three months' time the LIBOR rate is 5.9 percent, then the company would not enter the IRG and instead would negotiate a loan at the current LIBOR rate. This illustrates the beneficial aspect of the IRG over the FRA because companies have the option not to enter the contract when a favorable interest rate evolves at the time they wish to secure the loan.

Interest Rate Caps and Floors

An interest rate cap is a call option on interest rates that guarantees to its holder that a series of otherwise floating interest rates will not exceed a specified amount in the future. A borrower using an interest rate cap obtains protection against higher rates but can enjoy the lower borrowing cost if interest rates decrease. The most common example of caps in use involves a floating rate residential mortgage commonly referred to as an ARM (adjustable rate mortgage). An adjustable rate mortgage includes a cap that includes a series of options called caplets that corresponds to each of the reset dates of the loan. Should rates rise above the cap rate, the borrower does not incur the additional interest expense; rather the lender who sold the borrower the interest rate cap bears the additional interest expense.

Analogous to an interest rate cap, an interest rate floor is a portfolio of put options on interest rates. Each individual option constituting the floor is known as a floorlet. A floor ensures that a series of interest rates will not go below a minimum specified level or floor in the future. The same floating rate mortgage may also have a floor or minimum interest rate payment that protects the lender. In this case, the borrower would not benefit from interest rates that decline below the floor.

Collars

A collar is the simultaneous purchase of a cap and sale of a floor for the same expiration. This is a common strategy for reducing the cost of the premium to insure against an adverse movement in short term interest rates. The purchaser of the collar hedges against an upward swing in interest rates but simultaneously relinquishes the right to benefit from any significant fall in interest rates. Furthermore should rates fall below the level of the floor, the purchaser is obliged to pay the difference between the floor strike and the prevailing market rates. The premium obtained from the sale of the floor will in most cases only partially offset the cost of the cap. This cost can be reduced by raising the strike of the floor. When the premium of the floor exactly matches that of the cap, this is known as a zero-cost collar. Often collars are constructed to operationalize the users' view of where interest rates will be in the future. The magnitude of the difference between the strikes of the cap and floor will match users' views on volatility, with a large (small) difference corresponding to high (low) volatility.

Suppose Caribou Research, Inc. just completed an acquisition that will be financed with a bank loan. It secures a six-year, $500 million loan based on LIBOR pricing. The financing includes a 50 percent hedge requirement for a minimum term of three years. Caribou Research, Inc. and its lenders have agreed that the maximum LIBOR rate threshold to comfortably support cash flow projections is 8.50 percent. Therefore, the collar will have a cap set at 8.50 percent. Caribou wants to minimize the up-front cost associated with a hedge and also wants to retain some flexibility if LIBOR declines. So it wants the floor in the collar to be set at 6.5 percent. The collar reflects Caribou management's view that three-month LIBOR, currently at 7.50 percent, is headed higher over the foreseeable future but may eventually reverse course. Every three months, the debt will reprice based on the prevailing LIBOR rate. Simultaneously, any payments due under the collar will be determined. If LIBOR has risen above 8.50 percent, Caribou Research will be reimbursed for the difference. For example, if LIBOR is at 10.00 percent, Caribou Research will be credited for $(10.00\% - 8.50\%) = 1.50\%$. If LIBOR has declined below 6.50 percent to, say, 6.00 percent, Caribou Research will owe $(6.50\% - 6.00\%) = .50\%$ on the collar. If LIBOR is between 8.50 and 6.50 percent, no collar payments are due. In sum, LIBOR can fluctuate no higher than 8.50 percent and no lower than 6.50 percent on 50 percent of Caribou's debt over the next three years.

The benefits are that a collar reduces the cost of interest rate protection and protects against higher interest rates, and the user can sell the collar back to a bank at any time. The main disadvantage of a collar is that Caribou has to pay a certain minimum rate of interest, thus losing some of the possible benefit of lower interest rates.

Swaptions

A swaption is an option to enter into an interest rate swap on some future date. Swaptions often are used to hedge against future borrowings. There are two primary types of swaptions: payer's swaptions and receiver's swaptions. A payer's swaption gives buyers the right to pay the fixed rate on the swap, thus protecting them from raising rates. The receiver's swaption gives buyers the right to receive the fixed rate, thus protecting them from falling rates.

Swaptions are very similar to other options, but in this case, the underlying instrument is an interest rate swap. The expiry date is the date upon which the swaption may be exercised into the underlying swap, while the strike price is the fixed rate of the underlying swap. European swaptions can be exercised into the underlying swap or cash settled solely on the expiry date of the swaption. American swaptions are also available, and they come in two varieties. The first is a variable swaption, where the underlying swap has a fixed tenor, no matter when the swaption is exercised. The second is a wasting swaption, where the underlying swap has a fixed maturity date, so the tenor of the underlying swap becomes shorter the later the swap is exercised.

To illustrate the use of swaptions, consider the next situation. Caribou Research, Inc. knows it will finance a project by borrowing a lump sum of money in one year for three years by accessing its line of credit at Telluride Bank. Its line of credit is at a floating rate of LIBOR + 0.65 percent. Caribou Research faces the risk that interest rates may fluctuate over the next four years, so it prefers to borrow fixed. It considers two approaches. First, it can pay the fixed rate on a three-year swap (deferred one year) at a rate of 7 percent. In this case, the company locks in a fixed rate of 7 percent regardless of market rates in one year. If rates rise to 8 percent, the company effectively borrows at 7 percent under the swap. If rates fall to 6 percent, Caribou Research still borrows at 7 percent and cannot benefit from the lower spot rates. The second approach is to buy a one-year payer's swaption into a three-year 7 percent swaption for a swaption premium of 0.25 percent per year. Now if rates rise to 8 percent, Caribou Research exercises the swaption to fix at 7 percent. If rates fall to 6 percent, Caribou Research allows the option to expire worthless, and it borrows at 6 percent. The swaption allows the company to benefit from the falling rates whereas the deferred swap does not.

Mortgage Securitization Risk Management Using Interest Rate Derivatives

Mortgages are traditionally very illiquid assets. However, through the securitization process, mortgages are transformed into standardized instruments that are normally highly liquid. Mortgage lenders securitize mortgages by placing the loans in a trust and then securities representing shares of assets (mortgages) in the trust are sold to investors. The proceeds from the sale of these securities provide funding to purchase more loans and repeat the securitization process. Throughout the securitization process, lenders face interest rate risk, prepayment risk, and fallout. For purposes of this discussion, the pipeline extends from the rate lock to sale in to the secondary market, typically a period lasting between 60 to 90 days. Before a mortgage closes, falling interest rates can result in applicants withdrawing their loan applications; this is known as fallout. Interest rates falling after the closing can result in the homeowners prepaying and refinancing. In both instances, the mortgages drop out of the pipeline. Conversely, if interest rates rise before or after closing, homeowners will be more likely to stay in their current mortgage, and the number of mortgages in the pipeline remains the same. However, each mortgage will be worth less than par, and the lender cannot realize the full value from the sale of the paper. Clearly, interest rate risk directly generates challenges to pipeline managers, because both falling and rising interest rates can each have negative implications.

To manage these risks, pipeline managers use interest rate futures and options on futures to create an effective hedging strategy. The hedging manager must take into account the direct risk of changing interest rates on the value of the portfolio and the indirect risk of fallout. Besides changing interest rates, fallout also varies depending on the characteristics of the loans in the pipeline, such as loan type, purpose, source, and even which part of the country it comes from. Fallout can vary dramatically from branch to branch within the same lending institution. Therefore, hedging managers must estimate the level of fallout in their pipelines. This requires them to have enough quantitative resources to make reasonably good estimates of the likely effects of interest rate changes on fallout.

Let us look for a moment at how fallout affects a hedging strategy. In a rising rate environment, fallout can generate a significant increase in the potential loss of value in a mortgage lender's pipeline. As rates rise, loan values decline. In addition, the number of loans that close increases (i.e., as the rate locks gain in value to the consumer, fallout decreases). For this reason, lenders have to change their projections regarding the number of loans that will exit the other end of the pipeline and adjust the amount of the assets to be hedged.

This phenomenon is known as negative convexity, and it is a critically important source of risk for mortgage lenders. Basically, the lender's pipeline produces more loans at the worst possible time, and vice versa. When rates fall, fallout increases, and all the good loans that the lender thought it had locked in are not completed. All of this risk arises from one simple source: the put options that lenders offer to prospective borrowers. For this reason, using interest rate futures alone to hedge are not sufficient. Options on interest rate futures are a key element in hedging pipeline risk. The more interest rate sensitivity in the lender's pipeline, the more the lender will need to use options to hedge the potential fluctuations in pipeline value. Since lenders write a put when they offer the rate lock, they will tend to use puts, rather than calls, to offset that exposure.

Let us assume a pipeline manager has a \$40 million portfolio. After a thorough analysis of the mortgage portfolio, she decides to use forwards and futures to hedge part, but not all, of that pipeline, since she knows that some percentage of the loan commitments will drop out before closing. Additionally, she would use options on futures to hedge a portion of that exposure. As an example, the hedging strategy shown in Exhibit 37.2 is devised.

The manager estimates that \$30 million in loans will close; \$16 million for which she knows when they will close, and \$14 million for which she does not know the exact time when they will close. Many pipeline managers would hedge this exposure by simply selling forward their expected closing volume, either for cash or mortgage-backed securities. An alternative approach is to sell an equivalent amount of Treasury futures. As rates go up, the value of the short positions will rise and offset the losses on the loans in the pipeline. In any case, the manager uses a combination of forwards (for the portion of the portfolio she knows the date they

Exhibit 37.2 Hedging Strategy for a $40 Million Loan Portfolio

Amount	**Likely Outcome**	**Hedging Tool**
$16 million	Loans will close	Forwards
$14 million	Loans will close, unsure when	Futures
$8 million	Might close	Options on futures
$2 million	Will not close	No hedge

will be closed) and futures (for the portion of the portfolio for which she knows the loans will close but not exactly when). She also estimates that $2 million worth of the loans will not close for some reason or another and chooses not hedge this portion.

For the uncertain portion of the portfolio (due to potential fallout), suppose that the manager buys put options on futures to cover $8 million of the pipeline. If Treasury prices fall over the next few weeks or months and rates rise, those loans are likely to close, so she exercises the puts, which give her the right to sell Treasury futures at the price she fixed at the time of purchase. This allows the manager to lock in the futures at that price and install the hedge. The main expense—the hedge cost—is the premium paid to purchase the puts.

Now let us look at the opposite scenario. If Treasury prices rally and rates drop, the fallout ratio is going to rise, so the $8 million of loans that were covered with the put options drop out of the pipeline. Now there is no need for the hedge, so the manager decides against exercising the options, and they expire worthless. This hedging strategy exposes the manager to some loss, but the loss is limited to the premiums paid to purchase the options. Remember that if she had gone directly into the futures instead of using the options, the fallout would have required her to liquidate the hedge at a higher price than when purchased, and that would have been much more costly than the lost premiums.

The third scenario is that interest rates do not move at all. In this case, the manager buys the coverage by exercising the put options. In other words, she ends up with the same hedge profile as in the first scenario. The difference is that she loses the options premium and gains the peace of mind that comes from knowing that the hedging strategy is going to work in both rising and falling interest environments.

CONCLUSION

Interest rate derivatives come in a wide variety of standardized forms. They can be easily tailored to specific interest rate risk management needs of the user. This chapter has presented an overview of the types of instruments available and examples of their usefulness in managing interest rate risk.

REFERENCES

Bierwag, G. O. 1987. *Duration Analysis: Managing Interest Rate Risk.* Cambridge, MA: Ballinger Publishing.

Kolb, R. W., and J. A. Overdahl. 2007. *Futures, Options, and Swaps*. London: Blackwell Publishing.

SUGGESTED FURTHER READING

Cornyn, A., R. Klein, and J. Lederman. 1997. *Controlling & Managing Interest-Rate Risk*. New York: New York Institute of Finance.

James, T. 2006. *Interest Rate Derivatives: A Practical Guide to Applications, Pricing and Modeling*. London: Risk Books.

Obazee, P. O. 1997. *Interest Rate Derivatives: Analysis, Valuation and Market Applications*. New York: McGraw-Hill.

Fabozzi, F. J. 2002. *The Handbook of Financial Instruments*. Hoboken, NJ: John Wiley & Sons.

Kolb, R. W. Kolb, and J. A. Overdahl. 2007. *Futures, Options, and Swaps*. London: Blackwell Publishing.

Hull, J. 2008. *Options, Futures, and Other Derivatives*. Upper Saddle River, NJ: Prentice-Hall.

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

Jorion, P. 2007. *Financial Risk Manager Handbook*. Hoboken, NJ: John Wiley & Sons.

ABOUT THE AUTHOR

Steven Byers is currently an academic fellow at the U.S. Securities and Exchange Commission. He earned an International MBA from the University of Denver in 1992, a PhD in economics from Colorado State University in 1999, and a MSc in mathematical trading and finance from City University in London in 1998. Previously, Dr. Byers was the chief economist at the Public Company Accounting Oversight Board, where he led multiple fraud detection modeling projects and performed regulatory cost-benefit analysis. Prior to that, he worked at the Comptroller of the Currency in the Risk Analysis Division and at the Commodity Futures Trading Commission. He has authored published articles in peer-reviewed journals on topics related to general equilibrium modeling, regional economic modeling, and interest rate forecasting.

Index

KOLB SERIES IN FINANCE

Essential Perspectives

FINANCIAL DERIVATIVES

The Robert W. Kolb Series in Finance is an unparalleled source of information dedicated to the most important issues in modern finance. Each book focuses on a specific topic in the field of finance, and contains contributed chapters from both respected academics and experienced financial professionals. As part of the Robert W. Kolb Series in Finance, *Financial Derivatives* aims to provide a comprehensive understanding of financial derivatives and how you can prudently use them within the context of your underlying business activities.

For the public at large, financial derivatives have long been the most mysterious and least understood of all financial instruments. Through in-depth insights gleaned from years of financial experience, the contributors in this collection clearly explain what derivatives are without getting bogged down by the mathematics surrounding their pricing and valuation.

Financial Derivatives offers a broad overview of the different types of derivatives—futures, options, swaps, and structured products—while focusing on the principles that determine market prices. This comprehensive resource also provides a thorough introduction to financial derivatives and their importance to risk management in a corporate setting. Filled with in-depth analysis and examples, *Financial Derivatives* offers readers a wealth of knowledge on futures, options, swaps, financial engineering, and structured products.

ISBN 978-0-470-49910-8
59500
9 780470 499108