BARRA

United States Equity

Version 3 (E3)

RISK MODEL HANDBOOK

0111 O 02/98

Contents

Section II: US-E3 Model Details

About BARRA

In recent years the investment management industry has adjusted to continuing changes—theoretical advances, technological developments, and market growth. To address these challenges, investment managers and financial institutions require the most advanced and powerful analytical tools available.

A pioneer in risk management

As the leading provider of global investment decision tools, BARRA has responded to these industry changes by providing quantitative products and services that are both flexible and efficient. Since our founding in 1975, BARRA has been a leader in modern financial research and techniques.

Initially, our services focused on risk analysis in equity markets. Our U.S. Equity Model set a standard of accuracy that BARRA continues to follow. BARRA uses the best data available to develop econometric financial models. In turn, these models are the basis of software products designed to enhance portfolio performance through returns forecasting, risk analysis, portfolio construction, transaction cost analysis, and historical performance attribution.

In 1979, BARRA expanded into the fixed income area with the release of U.S. bond valuation and risk models. In the mid-1980s we developed a global tactical asset allocation system: The BARRA World Markets Model™. More recently, the Total Plan Risk™ approach was developed to provide multi-asset-class value-at-risk (VAR) analyses.

BARRA now has offices around the world and products that cover most of the world's traded securities. By 1997, our clients comprised approximately 1,200 financial institutions worldwide managing over $7 trillion in assets. They rely on BARRA's investment technology and consulting services to strengthen their financial analysis and investment decision-making.

Introduction

In this handbook

Section I: Theory contains a general discussion of equity risk and return, and the methods BARRA uses to model portfolio risk. Chapters 1 through 5 comprise this section.

Chapter 1. Why Risk is Important gives an overview of why financial professionals should care about risk.

Chapter 2. Defining Risk outlines the basic statistical concepts underlying risk analysis, and traces the history of equity risk theory.

Chapter 3. Modeling and Forecasting Risk discusses the application of multiple-factor modeling (MFM) to the equity risk analysis problem.

Chapter 4. Modern Portfolio Management and Risk relates the various types of active and passive equity management to the use of a risk model.

Chapter 5. BARRA Multiple-Factor Modeling details the process of creating and maintaining a BARRA equity MFM.

Section II: US-E3 Model Details discusses the construction of our third-generation U.S. equity risk model in depth. Chapters 6 through 12 and Appendices A through D comprise this section.

Chapter 6. Advantages of US-E3 Over US-E2 summarizes the reasons for updating US-E2 and the particular advances made with US-E3 over US-E2.

Chapter 7. The US-E3 Estimation Universe discusses the expansion of the US-E3 equity portfolio which is used to estimate the main parameters of the risk model. This was done to ensure a more accurate reflection of the investing activities of our clients.

Chapter 8. US-E3 Risk Indices and Descriptors describes differences from the US-E2 treatment of these factors and improvements that have been made in this part of the model.

Chapter 9. US-E3 Industries discusses the complete reclassification of the industry assignments and other enhancements intended to keep the model current over time. One of the least satisfactory aspects of US-E2 was the unchanging nature of the industry factor specification.

Chapter 10. Factor Return Estimation contains a brief description of US-E3's particular implementation of this process. A more general discussion is contained in Chapter 5.

Chapter 11. Estimating the Factor Covariance Matrix in US-E3 also briefly discusses only those aspects of this subject which are particular to US-E3, with a more general treatment detailed in Chapter 5.

Chapter 12. US-E3 Specific Risk Modeling also briefly discusses only those aspects of this subject which are particular to US-E3, with a more general treatment detailed in Chapter 5.

Appendix A: US-E3 Descriptor Definitions is the complete list of risk indices and their underlying data descriptors.

Appendix B: US-E3 Industries, Mini-Industries, Example Companies, and Codes contains the complete list of industries, their constituent "mini-industries," and selected example companies.

Appendix C: US-E3 Frequency Distributions for Predicted Beta, Specific Risk, Risk Indices contains graphical depictions of the distribution of these model outputs.

Appendix D: US-E3 Risk Index Factor Returns gives a full model history of the returns to the US-E3 risk indices.

Finally, the Glossary and Index are useful resources for clarifying terminology and enhancing the handbook's usefulness.

Further references

BARRA has a comprehensive collection of articles and other materials describing the models and their applications. To learn more about the topics contained in this handbook, consult the following references or our extensive Publications Bibliography, which is available from BARRA offices and from our Web site at *http://www.barra.com*.

Books

Andrew Rudd and Henry K. Clasing, *Modern Portfolio Theory: The Principles of Investment Management*, Orinda, CA, Andrew Rudd, 1988.

Richard C. Grinold and Ronald N. Kahn: *Active Portfolio Management: Quantitative Theory and Applications*, Probus Publishing, Chicago, IL, 1995.

1. Why Risk is Important

Superior investment performance is the product of careful attention to four elements:

- forming reasonable return expectations
- controlling risk so that the pursuit of opportunities remains tempered by prudence
- controlling costs so that investment profits are not dissipated in excessive or inefficient trading
- controlling and monitoring the total investment process to maintain a consistent investment program

These four elements are present in any investment management problem, be it a strategic asset allocation decision, an actively managed portfolio, or an index fund—managed bottom-up or top-down, via traditional or quantitative methods.

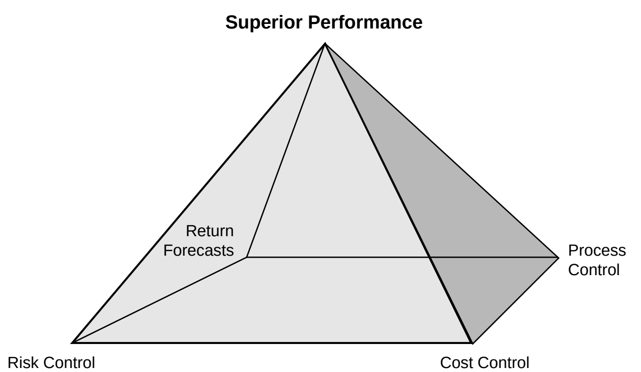


Figure 1-1
The Performance Pyramid

In a simpler view, return and risk are the protagonist and antagonist of investing. According to an old adage, the tradeoff between return and risk is the tradeoff between eating well and sleeping well. Clearly, risk doesn't just matter to quants!

Ignoring risk is hazardous to your portfolio. The optimal strategy ignoring risk places the entire portfolio in one stock. But no institutional investor follows this strategy. Hence risk considerations must impact every institutional portfolio. Unfortunately, they sometimes do not impact them enough.

We need not look far to find examples of financial disasters that arose through lack of sufficient risk control. The debacles of Orange County, Barings Bank, and the Piper Jaffray Institutional Government Income Fund all testify to the dangers of ignoring or poorly understanding risk.

But risk analysis is more than avoiding disasters—it can in fact enhance opportunities. Peter Bernstein has argued that a lack of understanding of risk holds back economic development.[1] Modern economic growth requires understanding risk.

What are the expected returns to a new venture? What are the risks? Do the returns outweigh the risks? Can I hedge the risks? In modern economies, the future is not beyond management, not simply subject to the whims of many gods. In fact, the period which marked the development of probability and statistics (during and after the Renaissance) also marked a time of profound growth in trade, exploration, and wealth. The ideas of risk management enabled the modern economic world, according to Bernstein. Risk analysis enhanced opportunities.

While Bernstein's argument may seem inspiring—though not of day-to-day relevance—in fact the goal of risk analysis is not to minimize risk but to properly weigh risk against return. Sometimes risk analysis leads to taking more risk.

The goal of risk analysis

Risk is important. It is a critical element of superior investment performance. Good risk analysis should provide not only a number—a quantification of risk—but insight, especially insight into the "Performance Pyramid."

We have illustrated superior performance as a three-dimensional object. A single risk number is only one-dimensional. So what do we mean by insight?

1. See Peter L. Bernstein, *Against the Gods: The Remarkable Story of Risk*, John Wiley & Sons, New York, 1996.

Risk analysis should uncover not just overall risk, but the largest and smallest bets in the portfolio. Do the largest bets correspond to the highest expected returns? They should. If they do not, the portfolio isn't properly balancing return and risk. Are the bets too large or too small? What is the "worst case" scenario? How will the portfolio compare to its benchmark?

Robust risk analysis can provide answers to all these questions as well as insight to all investors. In this volume we will discuss the history and current practice of equity risk modeling for single country markets. Other methods are used for different securities, such as bonds or currencies, and for different market structures, such as the global stock market. The underlying message is clear: The investor armed with superior methods of assessing and controlling risk possesses a significant competitive edge in modern capital markets.

2. Defining Risk

Some basic definitions

In an uncertain investment environment, investors bear risk. Risk is defined as the total dispersion or volatility of returns for a security or portfolio. Further, risk reflects uncertainty about the future.

We will define risk as the standard deviation of the return. Risk is an abstract concept. An economist considers risk to be expressed in a person's preferences. What is perceived as risky for one individual may not be risky for another.[1]

We need an operational and therefore universal and impersonal definition of risk. Institutional money managers are agents of pension fund trustees and other asset sponsors, who are themselves agents of the sponsoring organization and, ultimately, the beneficiaries of the fund. In that setting we cannot hope to have a personal view of risk.

We need a symmetric view of risk. Institutional money managers are judged relative to a benchmark or relative to their peers. The money manager suffers as much if he does not hold a stock and it goes up as if he held a larger than average amount of the stock and it goes down.

We need a flexible definition of risk. Our definition of risk should apply to individual stocks and to portfolios. We should be able to talk about realized risk in the past, and forecast risk over any future horizon.

The definition of risk that meets these criteria of being universal, symmetric, and flexible is the standard deviation of return.[2,3] If R_P is a portfolio's total return, then the portfolio's standard deviation of

1. There is a vast literature on this subject. The books of Arrow, Raiffa, and Borch are a good introduction.
2. An economist would call the standard deviation a measure of uncertainty rather than risk.
3. There is something of a debate currently over using measures of "downside" risk instead of volatility. Given the symmetric nature of active returns, U.S. institutional investors' avoidance of strategies like portfolio insurance which skew portfolio returns, and the practical limitations of analyzing large portfolios (>100 names), downside risk is inappropriate or irrelevant for active management. For a discussion, *see* Ronald Kahn and Dan Stefek, "Heat, Light, and Downside Risk," 1997.

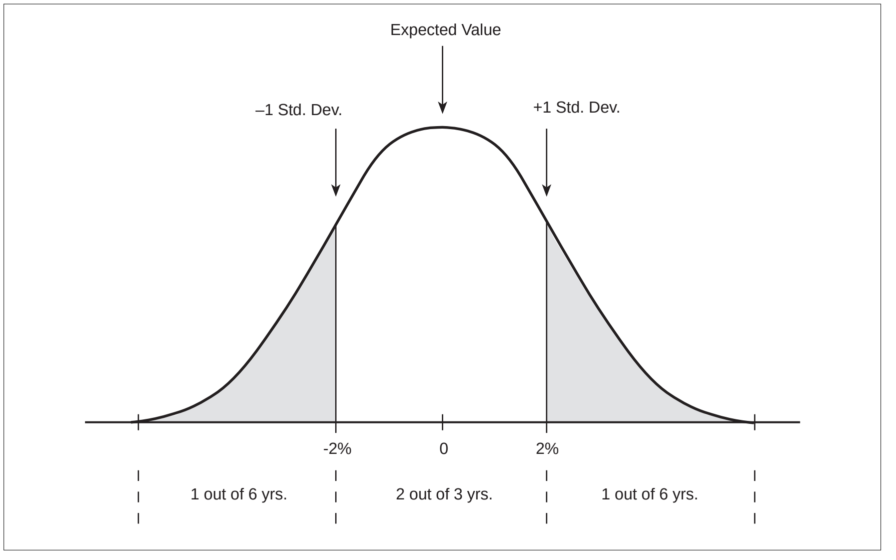
return is denoted by $\sigma_P \equiv Std[R_P]$. A portfolio's excess return r_P differs from the total return R_P by a constant R_F, so the risk of the excess return is equal to the risk of the total return. We will typically quote this risk, or standard deviation of return, on a percent per year basis. We will also occasionally refer to this quantity as volatility.[1]

The rough interpretation of standard deviation is that the return will be within one standard deviation of its expected value two-thirds of the time and within two standard deviations nineteen times out of twenty. Figure 2-1 graphically illustrates this fact.

Figure 2-1

Risk: The Dispersion of Returns

The standard deviation is a statistical measure of dispersion around an expected value—in this case, zero.



1. For a more detailed discussion of these concepts, please see Richard C. Grinold and Ronald N. Kahn, *Active Portfolio Management: Quantitative Theory and Applications*, Probus Publishing, Chicago, IL, 1995.

A related risk measure is variance, the standard deviation squared. The formulae are:

$$Std[\tilde{r}] = \sqrt{Var[\tilde{r}]} \qquad \text{(EQ 2-1)}$$

$$Var[\tilde{r}] = \mathrm{E}\left[(\tilde{r} - \bar{r})^2\right] \qquad \text{(EQ 2-2)}$$

where:

$\tilde{r}$ = return,

$\bar{r}$ = expected or mean return,

$Std[x]$ = standard deviation of x,

$Var[x]$ = variance of x, and

$\mathrm{E}[x]$ = expected value of x.

The standard deviation is the more common risk indicator since it is measured in the same units as return. Of course, if the standard deviation is known, the variance can be easily computed and vice versa. Other measures, including value-at-risk and shortfall risk, can be easily computed from the standard deviation.

An example

The standard deviation has some interesting characteristics. In particular it does *not* have the portfolio property. The standard deviation of a stock portfolio is not the weighted average of the standard deviations of the component stocks.

For example, suppose the correlation between the returns of Stocks 1 and 2 is ρ_{12}. If we have a portfolio of 50% Stock 1 and 50% Stock 2, then:

$$\sigma_P = \sqrt{(0.5 \cdot \sigma_1)^2 + (0.5 \cdot \sigma_2)^2 + 2 \cdot (0.5 \cdot \sigma_1)(0.5 \cdot \sigma_2) \cdot \rho_{12}} \qquad \text{(EQ 2-3)}$$

and $\sigma_P \le 0.5 \cdot \sigma_1 + 0.5 \cdot \sigma_2$ (EQ 2-4) with such equality being maintained only if the two stocks are perfectly correlated ($\rho_{12} = 1$). For risk, the whole is less than the sum of its parts. This is the key to portfolio diversification.

Figure 2-2

Risk Reduction Benefits of Diversification: A Two-Stock Example

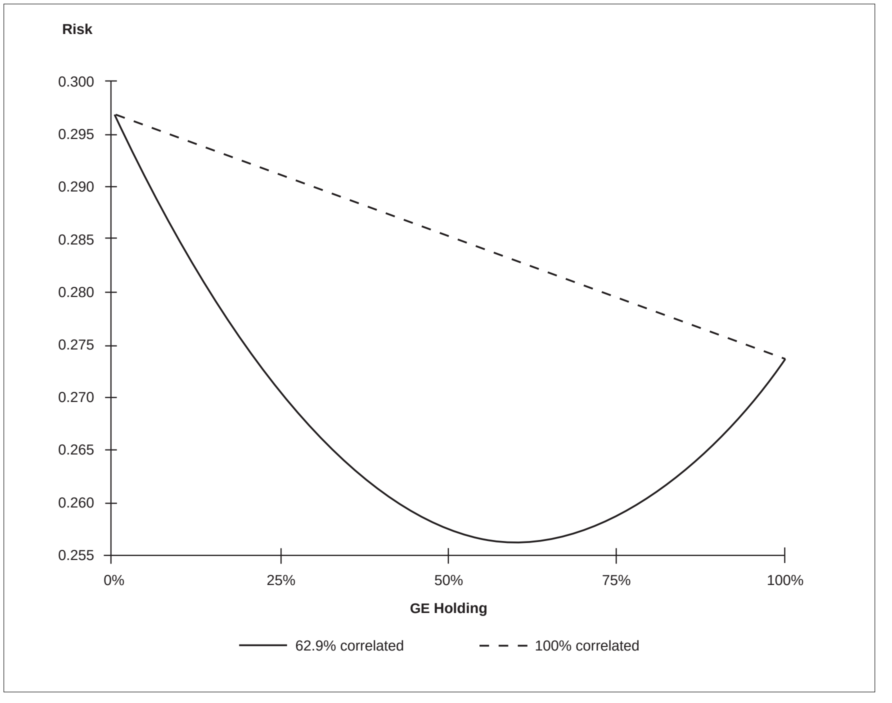


Figure 2-2 shows a simple example. The risk of a portfolio made up from IBM and General Electric is plotted against the fraction of GE stock in the portfolio. The curved line represents the risk of the portfolio; the straight line represents the risk that we would obtain if the returns on IBM and GE were perfectly correlated. The risk of GE is 27.4% per year; the risk of IBM is 29.7% per year; and the two returns are 62.9% correlated. The difference between the two lines is an indication of the benefit of diversification in reducing risk.

Risk reduction through diversification

We can see the power of diversification in another example. Given a portfolio of N stocks, each with risk σ and uncorrelated returns, the risk of an equal-weighted portfolio of these stocks will be:

$$\sigma_P = \frac{\sigma}{\sqrt{N}} \tag{EQ 2-5}$$

Note that the average risk is σ, while the portfolio risk is $\sigma/\sqrt{N}$.

For a more useful insight into diversification, assume now that the correlation between the returns of all pairs of stocks is equal to ρ. Then the risk of an equally weighted portfolio is:

$$\sigma_P = \sigma \cdot \sqrt{\frac{1 + \rho \cdot (N - 1)}{N}} \qquad \text{(EQ 2-6)}$$

In the limit that the portfolio contains a very large number of correlated stocks, this becomes:

$$\sigma_P \Rightarrow \sigma \cdot \sqrt{\rho} \qquad \text{(EQ 2-7)}$$

To get a feel for this, consider the example of an equal-weighted portfolio of the 20 Major Market Index constituent stocks. In December 1992, these stocks had an average risk of 27.8%, while the equal-weighted portfolio has a risk of 20.4%.[1] Equation 2-6 then implies an average correlation between these stocks of 0.52.

Risks don't add across stocks and risks don't add across time. However, variance will add across time if the returns in one interval of time are uncorrelated with the returns in other intervals of time. The assumption is that returns are uncorrelated from period to period. The correlation of returns across time (called *autocorrelation*) is close to zero for most asset classes. This means that variances will grow with the length of the forecast horizon and the risk will grow with the square root of the forecast horizon. Thus, a 5% annual active risk is equivalent to a 2.5% active risk over the first quarter or a 10% active risk over four years. Notice that the variance over the quarter, year, and four-year horizon (6.25, 25, and 100) remains proportional to the length of the horizon.

We use this relationship every time we "annualize" risk—i.e., standardize our risk numbers to an annual period. If we examine monthly returns to a stock and observe a monthly return standard deviation of $\sigma_{monthly}$, we convert this to annual risk according to:

$$\sigma_{annual} = \sqrt{12} \cdot \sigma_{monthly} \qquad \text{(EQ 2-8)}$$

1. These are predicted volatilities from BARRA's U.S. Equity Model.

Drawbacks of simple risk calculations

The mathematical calculation of risk using standard deviation of returns is therefore straightforward and can be extended to any number of securities. However, this approach suffers from several drawbacks:

- Estimating a robust covariance matrix of returns requires data for as many periods as we have securities to analyze. For large markets, such as the U.S. stock market, these long returns histories may simply not be available.
- Estimation error may occur in any one period due to spurious asset correlations that are unlikely to repeat in a systematic fashion.
- A simple covariance matrix of returns offers little in the way of economic analysis. In other words, it is largely a “black box” approach with little intuitive basis or forecastability.

For all these reasons, financial economists have sought for many years to model investment risk in more nuanced ways. We will now turn to a brief history of these efforts.

Evolution of concepts

The development of equity risk concepts has evolved from the modest and unscientific guesswork of early investment theory to the quantitative analysis and technical sophistication of modern financial tools.

“Buy a stock. If it goes up, sell it. If it goes down, don’t buy it.”
Will Rogers, 1931

Before the 1950s, there was no concept of systematic, or market-related, return. Return was a rise in the value of a stock and risk was a drop in the value of a stock. The investor’s primary investment tools were intuition and insightful financial analysis. Portfolio selection was simply an act of assembling a group of “good” stocks.

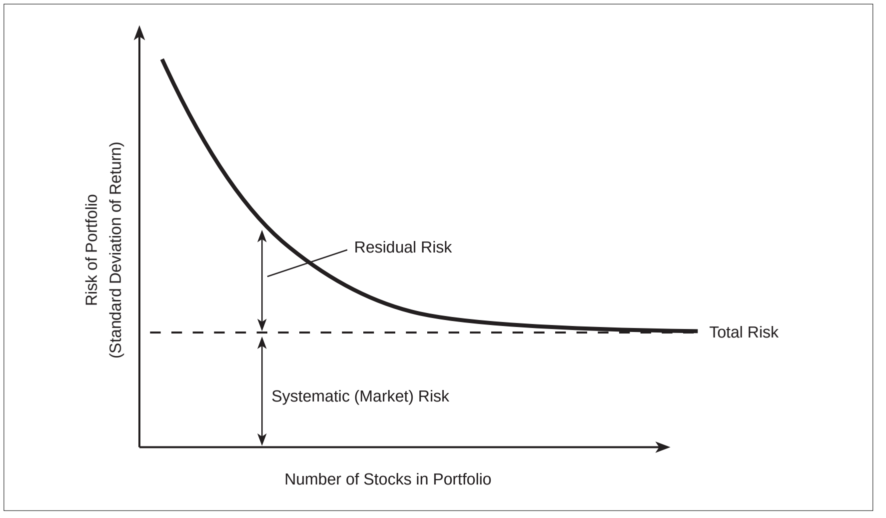


Figure 2-3

Diversification and Risk

As a portfolio manager increases the number of stocks in a portfolio, residual—or non-market-related—risk is diversified. Market risk is undiversifiable.

Financial theorists became more scientific and statistical in the early 1950s. Harry Markowitz was the first to quantify risk (as standard deviation) and diversification. He showed precisely how the risk of the portfolio was less than the risk of its components. In the late 1950s, Breiman and Kelly derived mathematically the peril of ignoring risk. They showed that a strategy that explicitly accounted for risk outperformed all other strategies in the long run.[1]

"Diversification is good."
Harry Markowitz, 1952

We now know how diversification affects risk exposures. It averages factor-related risk, such as industry exposures, and significantly reduces security-specific risk. However, diversification does not eliminate all risk because stocks tend to move up and down together with the market. Therefore, systematic, or market, risk cannot be eliminated by diversification.

Figure 2-3 shows the balance between residual risk and market risk changing as the number of different stocks in a portfolio rises. At a certain portfolio size, all residual risk is effectively removed, leaving only market risk.

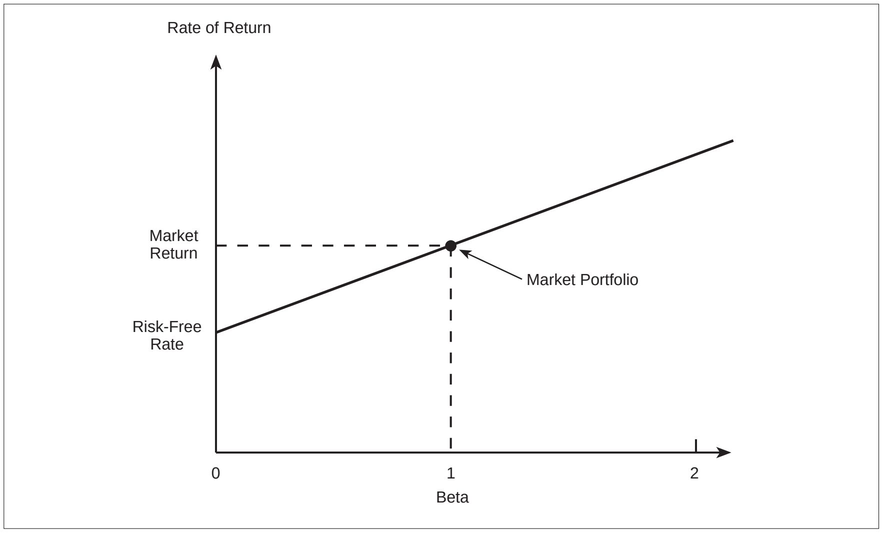
As investment managers became more knowledgeable, there was a push to identify the conceptual basis underlying the concepts of risk, diversification, and returns. The Capital Asset Pricing Model (CAPM) was one approach that described the equilibrium relationship between return and systematic risk. William Sharpe earned the Nobel Prize in Economics for his development of the CAPM.

1. See, for example, Leo Breiman, "Investment Policies for Expanding Businesses Optimal in a Long-Run Sense," *Naval Research Logistics Quarterly*, Vol. 7, No. 4, December 1960, pp. 647–651.

Figure 2-4

The Capital Asset Pricing Model

The Capital Asset Pricing Model asserts that the expected excess return on securities is proportional to their systematic risk coefficient, or beta. The market portfolio is characterized by a beta of unity.



The central premise of CAPM is that, on average, investors are not compensated for taking on residual risk. CAPM asserts that the expected residual return is zero while the expected systematic return is greater than zero and linear (*see* Figure 2-4).

The measure of portfolio exposure to systematic risk is called *beta* (β). Beta is the relative volatility or sensitivity of a security or portfolio to market moves. More simply, beta is the numerical value of an asset's systematic risk. Returns, and hence risk premiums, for any stock or portfolio will be related to beta, the exposure to undiversifiable systematic risk. Equation 2-9 states this linear relationship.

$$\mathrm{E}[\tilde{r}_i] - r_F = \beta_i \mathrm{E}[\tilde{r}_M - r_F] \tag{EQ 2-9}$$

where:

"Only undiversifiable risk should earn a premium."

William F. Sharpe, 1964

Capital Asset Pricing Model

$\tilde{r}_i$ = return on asset i,

r_F = risk-free rate of return,

$\tilde{r}_M$ = return on market portfolio, and

$$\beta_i = \frac{Cov[\tilde{r}_i, \tilde{r}_M]}{Var[\tilde{r}_M]}$$

The CAPM is a model of return. Underlying it are equilibrium arguments and the view that the market is efficient because it is the portfolio that every investor on average owns. The CAPM does not require that residual returns be uncorrelated. But it did inspire Sharpe to suggest a one-factor risk model that does assume uncorrelated residual returns. This model has the advantage of simplicity. It is quite useful for back of the envelope calculations. But it ignores the risk that arises from common factor sources, such as industries, capitalization, and yield.

By the 1970s, the investment community recognized that assets with similar characteristics tend to behave in similar ways. This notion is captured in the Arbitrage Pricing Theory (APT). APT asserts that security and portfolio expected returns are linearly related to the expected returns of an unknown number of underlying systematic factors.

"The arbitrage model was proposed as an alternative to the mean variance capital asset pricing model."

Stephen A. Ross, 1976

Arbitrage Pricing Theory

The focus of APT is on forecasting returns. Instead of equilibrium arguments, Ross and others used arbitrage arguments to assert that expected specific returns are zero, but expected common factor returns (including the market and other factors) need not be zero. Just like the CAPM, APT inspired a class of risk models: the multiple-factor model (MFM). The basic premise of an MFM is that many influences act on the volatility of a stock, and these influences are often common across many stocks. A properly constructed MFM is able to produce risk analyses with more accuracy and intuition than a simple covariance matrix of security returns or the CAPM.

"Since the factors can represent the components of return as seen by the financial analyst, the multiple-factor model is a natural representation of the real environment."

Barr Rosenberg, 1974

Multiple Factor Models

Multifactor models of security market returns can be divided into three types: macroeconomic, fundamental, and statistical factor models. Macroeconomic factor models use observable economic time series, such as inflation and interest rates, as measures of the pervasive shocks to security returns. Fundamental factor models use the returns to portfolios associated with observed security attributes such as dividend yield, the book-to-market ratio, and industry membership. Statistical factor models derive their factors from factor analysis of the covariance matrix of security returns. BARRA research has confirmed that of these three, fundamental factor models outperform the other two types in terms of explanatory power.[1]

We now turn to a discussion of fundamental MFMs in more detail.

1. Gregory Connor, "The Three Types of Factor Models: A Comparison of Their Explanatory Power," *Financial Analysts Journal*, May/June 1995.

3. Modeling and Forecasting Risk

Through the years, theoretical approaches to investment analysis have become increasingly sophisticated. With more advanced concepts of risk and return, investment portfolio models have changed to reflect this growing complexity. The multiple-factor model (MFM) has evolved as a helpful tool for analyzing portfolio risk.

What are MFMs?

Multiple-factor models (MFMs) are formal statements about the relationships among security returns in a portfolio. The basic premise of MFMs is that similar stocks should display similar returns. This "similarity" is defined in terms of ratios, descriptors, and asset attributes which are based on market information, such as price and volume, or fundamental data derived from a company's balance sheet and income statement.

MFMs identify common factors and determine return sensitivity to these factors. The resulting risk model incorporates the weighted sum of common factor return and specific return. The risk profile will respond immediately to changes in fundamental information.

How do MFMs work?

We derive MFMs from securities patterns observed over time. The difficult steps are pinpointing these patterns and then identifying them with asset factors that investors can understand. Asset factors are characteristics related to securities price movements, such as industry membership, capitalization, and volatility.

Once model factors are chosen and assigned to individual assets in the proper proportions, cross-sectional regressions are performed to determine the returns to each factor over the relevant time period. This allows the model to be responsive to market changes in a timely fashion.

Risk calculation is the final step in constructing a sound and useful model. Variances, covariances, and correlations among factors are estimated and weighted. We then use these calculations to describe the risk exposure of a portfolio.

Investors rely on risk exposure calculations to determine stock selection, portfolio construction, and other investment strategies. They base their decisions on information gleaned from MFM analysis as well as their risk preferences and other information they possess.

Advantages of MFMs

There are several advantages to using MFMs for security and portfolio analysis.

- MFMs offer a more thorough breakdown of risk and, therefore, a more complete analysis of risk exposure than other methods such as simple CAPM approaches.

- Because economic logic is used in their development, MFMs are not limited by purely historical analysis.

- MFMs are robust investment tools that can withstand outliers.

- As the economy and individual firms change, MFMs adapt to reflect changing asset characteristics.

- MFMs isolate the impact of individual factors, providing segmented analysis for better informed investment decisions.

- From an applications viewpoint, MFMs are realistic, tractable, and understandable to investors.

- Lastly, MFMs are flexible models allowing for a wide range of investor preferences and judgment.

Of course, MFMs have their limitations. They predict much, but not all, of portfolio risk. In addition, a model does not offer stock recommendations; investors must make their own strategy choices.

A simple MFM

To illustrate the power of MFMs, let's begin with a simple example.

Accurate characterization of portfolio risk requires an accurate estimate of the covariance matrix of security returns. A relatively simple way to estimate this covariance matrix is to use the history of security returns to compute each variance and covariance. This approach, however, suffers from two major drawbacks:

- Estimating a covariance matrix for, say, 3,000 stocks requires data for at least 3,000 periods. With monthly or weekly estimation horizons, such a long history may simply not exist.

- It is subject to estimation error: in any period, two stocks such as Weyerhaeuser and Ford may show very high correlation—higher than, say, GM and Ford. Our intuition suggests that the correlation between GM and Ford should be higher because they are in the same line of business. The simple method of estimating the covariance matrix does not capture our intuition.

This intuition, however, points to an alternative method for estimating the covariance matrix. Our feeling that GM and Ford should be more highly correlated than Weyerhaeuser and Ford comes from Ford and GM being in the same industry. Taking this further, we can argue that firms with similar characteristics, such as their line of business, should have returns that behave similarly. For example, Weyerhaeuser, Ford, and GM will all have a common component in their returns because they would all be affected by news that affects the stock market as a whole. The effects of such news may be captured by a stock market component in each stock's return. This common component may be the (weighted) average return to all U.S. stocks. The degree to which each of the three stocks responds to this stock market component depends on the sensitivity of each stock to the stock market component.

Additionally, we would expect GM and Ford to respond to news affecting the automobile industry, whereas we would expect Weyerhaeuser to respond to news affecting the forest and paper products industry. The effects of such news may be captured by the average returns of stocks in the auto industry and the forest and paper products industry. There are, however, events that affect one stock without affecting the others. For example, a defect in the brake system of GM cars, that forces a recall and replacement of the system, will likely have a negative impact on GM's stock price. This event, however, will most likely leave Weyerhaeuser and Ford stock prices unaltered.

These arguments lead us to the following representation for returns:

$$\tilde{r}_{GM} = \mathrm{E}[\tilde{r}_{GM}] + \beta_{GM} \cdot [\tilde{r}_M - \mathrm{E}[\tilde{r}_M]] + 1 \cdot [\tilde{r}_{AUTO} - \mathrm{E}[\tilde{r}_{AUTO}]] + 0 \cdot [\tilde{r}_{FP} - \mathrm{E}[\tilde{r}_{FP}]] + u_{GM} \tag{EQ 3-1}$$

where:

$\tilde{r}_{GM}$ = GM's realized return,

$\tilde{r}_M$ = the realized average stock market return,

$\tilde{r}_{AUTO}$ = realized average return to automobile stocks,

$\tilde{r}_{FP}$ = the realized average return to forest and paper products stocks,

E[.] = expectations,

β_{GM} = GM's sensitivity to stock market returns, and

u_{GM} = the effect of GM specific news on GM returns.

This equation simply states that GM's realized return consists of an expected component and an unexpected component. The unexpected component depends on any unexpected events that affect stock returns in general $[\tilde{r}_M - \mathrm{E}[\tilde{r}_M]]$, any unexpected events that affect the auto industry $[\tilde{r}_{AUTO} - \mathrm{E}[\tilde{r}_{AUTO}]]$, and any unexpected events that affect GM alone (u_{GM}). Similar equations may be written for Ford and Weyerhaeuser.

The sources of variation in GM's stock returns, thus, are variations in stock returns in general, variations in auto industry returns, and any variations that are specific to GM. Moreover, GM and Ford returns are likely to move together because both are exposed to stock market risk and auto industry risk. Weyerhaeuser and GM, and Weyerhaeuser and Ford, on the other hand, are likely to move together to a lesser degree because the only common component in their returns is the market return. Some additional correlation would arise, however, because auto and forest and paper products industry returns may exhibit some correlation.

By beginning with our intuition about the sources of co-movement in security returns, we have made substantial progress in estimating the covariance matrix of security returns. What we need now is the covariance matrix of common sources in security returns, the variances of security specific returns, and estimates of the sensitivity of security returns to the common sources of variation in their returns. Because the common sources of risk are likely to be much fewer than the number of securities, we need to estimate a much smaller covariance matrix and hence a smaller history of returns is required. Moreover, because similar stocks are going to have larger sensitivities to similar common sources of risk, similar stocks will be more highly correlated than dissimilar stocks: our estimated correlation for GM and Ford will be larger than that for Ford and Weyerhaeuser.

The decomposition of security returns into common and specific sources of return is, in fact, a multiple-factor model of security returns. We now turn to a generalized discussion of this process for many factors.

Model mathematics

MFMs build on single-factor models by including and describing the interrelationships among factors. For single-factor models, the equation that describes the excess rate of return is:

$$\tilde{r}_j = X_j \tilde{f} + \tilde{u}_j \qquad \text{(EQ 3-2)}$$

where:

$\tilde{r}_j$ = total excess return over the risk-free rate,

X_j = sensitivity of security j to the factor,

$\tilde{f}$ = rate of return on the factor, and

$\tilde{u}_j$ = nonfactor (specific) return on security j.

We can expand this model to include K factors. The total excess return equation for a multiple-factor model becomes:

$$\tilde{r}_j = \sum_{k=1}^{K} X_{jk}\,\tilde{f}_k + \tilde{u}_j \tag{EQ 3-3}$$

where:

X_{jk} = risk exposure of security j to factor k, and

$\tilde{f}_k$ = rate of return on factor k.

Note that when K=1, the MFM equation reduces to the earlier single-factor version. For example, the CAPM is a single-factor model in which the "market" return is the only relevant factor.

When a portfolio consists of only one security, Equation 3-3 describes its excess return. But most portfolios comprise many securities, each representing a proportion, or weight, of the total portfolio. When weights $h_{P1}, h_{P2},\ldots,h_{PN}$ reflect the proportions of N securities in portfolio P, we express the excess return in the following equation:

$$\tilde{r}_P = \sum_{k=1}^{K} X_{Pk}\,\tilde{f}_k + \sum_{j=1}^{N} h_{Pj}\,\tilde{u}_j \tag{EQ 3-4}$$

where:

$$X_{Pk} = \sum_{j=1}^{N} h_{Pj}\,X_{jk}$$

This equation includes the risk from all sources and lays the groundwork for further MFM analysis.

Risk prediction with MFMs

Investors look at the variance of their total portfolios to provide a comprehensive assessment of risk. To calculate the variance of a portfolio, you need to calculate the covariances of all the constituent components.

Without the framework of a multiple-factor model, estimating the covariance of each asset with every other asset is computationally burdensome and subject to significant estimation errors. For example, using an estimation universe of 1,400 assets, there are 980,700 covariances and variances to calculate (*see* Figure 3-1).

$$V(i, j) = \text{Covariance}\,[r(\tilde{i}),\ r(\tilde{j})]$$

where $V(i, j)$ = asset covariance matrix, and
i, j = individual stocks.

$$V = \begin{bmatrix} V(1,1) & V(1,2) & \cdots & V(1,N) \\ V(2,1) & V(2,2) & \cdots & V(2,N) \\ V(3,1) & V(3,2) & \cdots & V(3,N) \\ \vdots & \vdots & & \vdots \\ V(N,1) & V(N,2) & \cdots & V(N,N) \end{bmatrix}$$

Figure 3-1
Asset Covariance Matrix
For N=1,400 assets, there are 980,700 covariances and variances to estimate.

An MFM simplifies these calculations dramatically. This results from replacing individual company profiles with categories defined by common characteristics (factors). Since the specific risk is assumed to be uncorrelated among the assets, only the factor variances and covariances need to be calculated during model estimation (*see* Figure 3-2).

$$\tilde{r} = X\tilde{f} + \tilde{u}$$

where $\tilde{r}$ = vector of excess returns,
X = exposure matrix,
$\tilde{f}$ = vector of factor returns, and
$\tilde{u}$ = vector of specific returns.

$$\begin{bmatrix} \tilde{r}(1) \\ \tilde{r}(2) \\ \vdots \\ \tilde{r}(N) \end{bmatrix} = \begin{bmatrix} X(1,1) & X(1,2) & \cdots & X(1,K) \\ X(2,1) & X(2,2) & \cdots & X(2,K) \\ \vdots & \vdots & & \vdots \\ X(N,1) & X(N,2) & \cdots & X(N,K) \end{bmatrix} \begin{bmatrix} \tilde{f}(1) \\ \tilde{f}(2) \\ \vdots \\ \tilde{f}(K) \end{bmatrix} + \begin{bmatrix} \tilde{u}(1) \\ \tilde{u}(2) \\ \vdots \\ \tilde{u}(N) \end{bmatrix}$$

Figure 3-2
Factor Return Calculation
Using an MFM greatly simplifies the estimation process. Figure 3-2 depicts the multiple-factor model in matrix terms.

By using a multiple-factor model, we significantly reduce the number of calculations. For example, in the U.S. Equity Model (US-E3), 65 factors capture the risk characteristics of equities. This reduces the number of covariance and variance calculations to 2,145 (*see* Figure 3-3). Moreover, since there are fewer parameters to determine, they can be estimated with greater precision.

Figure 3-3

Factor Covariance Matrix

For K=65 factors, there are 2,145 covariances and variances to estimate. Quadrant I includes the covariances of risk indices with each other; quadrants II and III are mirror images of each other, showing the covariances of risk indices with industries; and quadrant IV includes covariances of industries with each other.

$$F(k, m) = \text{Covariance}\left[\tilde{f}(k), \tilde{f}(m)\right]$$

where $F(k, m)$ = factor covariance matrix, and k, m = common factors.

$$F = \left[\begin{array}{ccc:ccc} F(1,1) & \ldots & F(1,13) & F(1,14) & \ldots & F(1,65) \\ \vdots & \mathrm{I} & \vdots & \vdots & \mathrm{II} & \vdots \\ F(13,1) & \ldots & F(13,13) & F(13,14) & \ldots & F(13,65) \\ \hdashline F(14,1) & \ldots & F(14,13) & F(14,14) & \ldots & F(14,65) \\ \vdots & \mathrm{III} & \vdots & \vdots & \mathrm{IV} & \vdots \\ F(65,1) & \ldots & F(65,13) & F(65,14) & \ldots & F(65,65) \end{array}\right]$$

We can easily derive the matrix algebra calculations that support and link the above diagrams by using an MFM. From Figure 3-2, we start with the MFM equation:

$$\tilde{r}_i = X\tilde{f} + \tilde{u} \qquad \text{(EQ 3-5)}$$

where:

$\tilde{r}_i$ =excess return on asset i,

X =exposure coefficient on the factor,

$\tilde{f}$ =factor return, and

$\tilde{u}$ =specific return.

Substituting this relation in the basic equation, we find that:

$$Risk = Var\left(\tilde{r}_j\right) \qquad \text{(EQ 3-6)}$$

$$= Var\left(X\tilde{f} + \tilde{u}\right) \qquad \text{(EQ 3-7)}$$

Using the matrix algebra formula for variance, the risk equation becomes:

$$Risk = XFX^T + \Delta \tag{EQ 3-8}$$

where:

X = exposure matrix of companies upon factors,

F = covariance matrix of factors,

X^T = transpose of X matrix, and

Δ = diagonal matrix of specific risk variances.

This is the basic equation that defines the matrix calculations used in risk analysis in the BARRA equity models.

4. Modern Portfolio Management and Risk

In the previous chapters we observed that risk modeling is essential to successful portfolio management and that the standard deviation of returns is the best numerical risk measure. We also traced the evolution of risk concepts from portfolio standard deviation of security returns, through the CAPM and APT, to the current application of multiple-factor models (MFMs) in the portfolio management problem. In this chapter we will briefly explore the components of portfolio management and how BARRA MFMs can assist the manager at various points. In the next chapter a detailed description of BARRA MFMs will complete the theoretical portion of this handbook.

Portfolio management—two types

An equity portfolio manager must choose between two management methods: passive or active. BARRA MFMs are used to facilitate both methods.

Passive management

Passive management is an outgrowth of CAPM logic. In its broadest sense, passive management refers to any management strategy that does not rely on the possession of superior information. More specifically, disclosure of a passive investment strategy offers no competitive information that would undermine the strategy's validity.

One type of passive management is indexing, tracking the performance of a particular index. An example is the "buy-and-hold" philosophy which exposes the portfolio only to systematic risk. The second form of passive management is constructing a portfolio to match prespecified attributes or constraints. The portfolio may be yield-biased with a selected beta or match an index within certain parameters. This is often called *enhanced indexing*.

Benchmark

A *benchmark* is the standard of comparison for investment performance and risk analysis. It is widely used to evaluate and track performance of investment managers. The benchmark is also known as the *normal portfolio*—that is, the asset basket a manager would hold in the absence of any judgmental information. It reflects the manager's particular style and biases.

Tracking Error

Tracking error is a measure of risk exposure. It is the annualized standard deviation of the difference between portfolio return and benchmark return.

Because it provides a relative measure of risk exposure, tracking error is a useful evaluation tool, particularly for passive portfolios. Moreover, it offers relevant performance comparisons because the benchmark is selected based on portfolio characteristics and investor objectives.

Passive management procedures are distinguished by the following attributes:

- They exclude any transactions in response to judgments about security valuations and the market as a whole.
- They contain relatively minimal residual risk with respect to the benchmark or index.
- They often involve industry or sector weighting.

BARRA MFMs facilitate passive management by providing robust portfolio risk estimates versus passive benchmarks. The indexed portfolio can be readily compared with the benchmark to determine the magnitude of active risk (or tracking error) and its composition. Corrective action can be taken based on individual holding analysis which reveals those securities contributing the most active risk. Optimizing software can also be used to automate the rebalancing process.

Active management

Alpha

Alpha (α) generally refers to the expected exceptional return of a portfolio, factor, or individual asset. The use of alphas is a distinction of active management. They indicate that a manager believes a portion of expected return is attributable to particular factors.

Historical alpha is the difference between actual performance and the performance of a diversified market portfolio with identical systematic risk over the same period. *Judgmental*, or *predicted*, *alpha* is the expected value of subsequent extraordinary return based on a return forecast.

Active management refers to investment strategies designed to increase return by using superior information. In other words, the active manager seeks to profit from information that would lose its value if all market participants interpreted it in the same way. If, for example, an investment manager observed that securities with certain characteristics performed better (or worse) than expected, the manager could hold a larger (or smaller) proportion of that security to increase the subsequent value of the portfolio.

By following active management strategies, investors can add value to their portfolio if they predict returns better than the consensus expectations of the market. Information is obtained through ongoing research to forecast such things as yield curve changes, factor and industry returns, and transitory mispricing. At any given time, portfolio construction should reflect the tradeoff between risk and return—that is, any contribution to risk should be offset by the contribution to reward.

There are several basic types of active investment strategies. They include *market timing*, *sectoral emphasis*, and *stock selection*.

Market timing is the process of altering market risk exposure based on short-term forecasts in order to earn superior returns. The manager seeks to sell before the market goes down and buy before the market goes up. However, this strategy increases the variability in the portfolio beta, inducing increased systematic risk through time. BARRA MFMs assist market timing by giving the investor a robust beta estimate for any portfolio and indicating the most efficient way to take on or reduce market risk, including the use of futures.

The second type of active management is *sectoral emphasis*. Sectoral emphasis can be thought of as a combination of the other active strategies. It is both factor timing and a broad version of stock selection. The manager attempts to increase residual return through manipulating common factor exposures. For example, the manager can bet on an industry of high-yield stocks. Because several sectors can be emphasized at any given time, diversification is possible. BARRA MFMs possess detailed industry and risk index exposure information that can be utilized for any combination of sectoral tilts.

Lastly, *stock selection* is a portfolio allocation strategy based on picking mispriced stocks. It uses security alphas to identify over- and undervalued stocks. The manager can then adjust the asset proportions in the portfolio to maximize specific return. These active holdings, in both positive and negative directions, increase residual risk and portfolio alpha. The primary objective of this strategy is to manage asset holdings so that any change in incremental risk is compensated by a comparable change in return. BARRA MFMs facilitate stock selection by extending the risk model down to the individual equity level.

Figure 4-1 illustrates the typical prevalence of these various types of risk in a single stock, a small portfolio, and a multiple-portfolio situation. In each case, the manager's goal is to earn a superior return with minimum risk. The use of a multiple-factor model permits the manager to pursue these active management strategies with maximum precision.

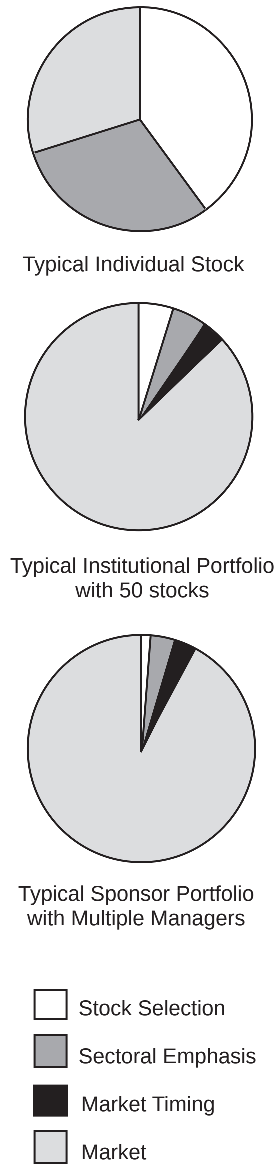


Figure 4-1

Diversification and Risk

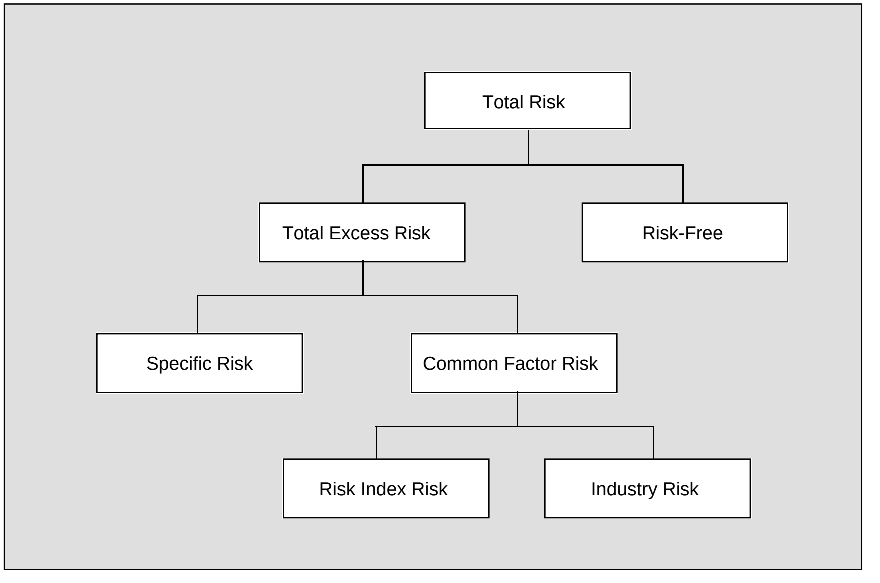
Market risk grows as a proportion of total risk as portfolio size increases.

Decomposing risk

BARRA's equity models isolate components of risk based on correlations across securities sharing similar characteristics. There are several ways to break down a portfolio's risk. BARRA uses any of four decompositions of risk, each reflecting a different perspective on portfolio management. These four decompositions are used in different ways by active and passive managers to provide insight and enhance performance.

Total Risk Decomposition

Figure 4-2
Total Risk Decomposition



The simplest risk decomposition, **Total Risk** (*see* Figure 4-2), is a basic breakdown into specific and common factor risk. There is no concept of a market, or systematic, portfolio. The risk is attributed purely to common factor and security-specific influences.

❖ *Common factor risk* is portfolio risk that arises from assets' exposures to common factors, such as capitalization and industries.

❖ *Specific risk* is unique to a particular company and thus is uncorrelated with the specific risk of other assets. For a portfolio, specific risk is the weighted sum of all the holdings' specific risk values.

This risk decomposition is useful for managers who wish to minimize total risk, or for managers such as hedge funds which may be pursuing market-neutral or other long/short strategies.

Systematic-Residual Risk Decomposition

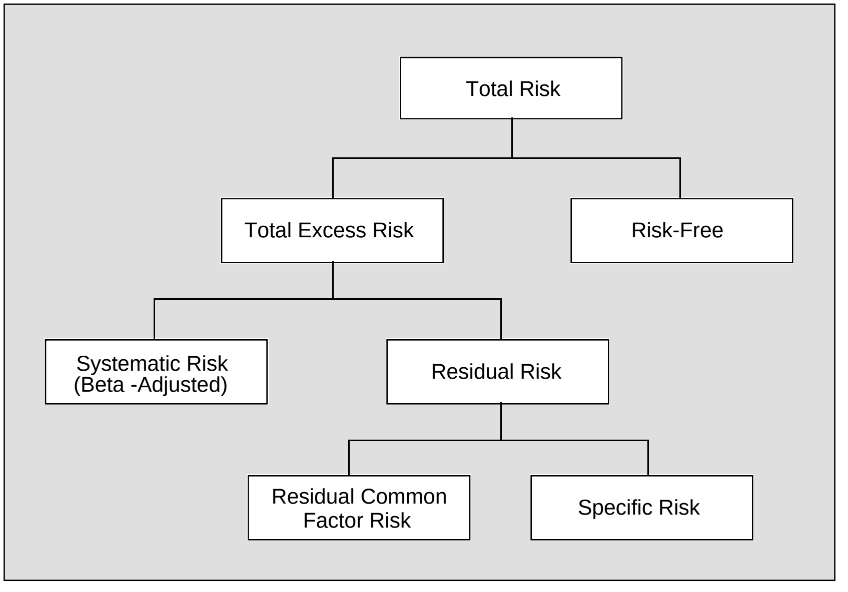


Figure 4-3
Systematic-Residual Risk Decomposition

This decomposition introduces the market into risk analysis (*see* Figure 4-3). This perspective partitions risk into the familiar categories of systematic (market) and residual risk.

- *Systematic risk* is the component of risk associated with the market portfolio. It is linked to the portfolio beta, which is the weighted average of the portfolio's asset betas.

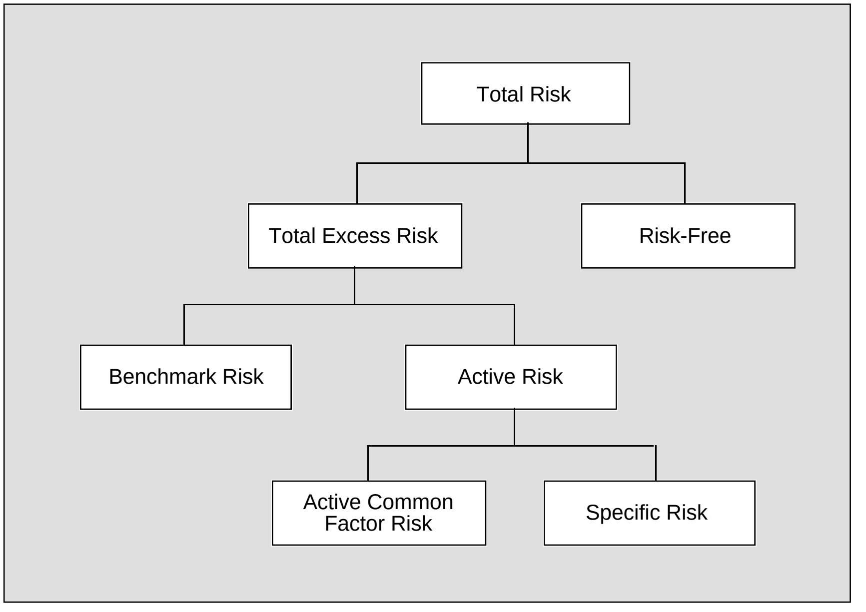
- *Residual risk* is the component of risk uncorrelated with the market portfolio. It is further divided into specific and common factor sources. Residual risk can be diversified to a negligible level.

This type of risk decomposition is the Capital Asset Pricing Model (CAPM) approach. With this approach, you can compare your managed portfolio against a broad-based market portfolio. A benchmark portfolio never comes into play. Risk is partitioned into *residual* and *systematic* components, and residual risk is further divided into specific and common factor sources.

This risk decomposition would be most useful to market timers or other managers who "tilt" away from the market portfolio on an opportunistic basis.

Active Risk Decomposition

Figure 4-4
Active Risk Decomposition



In this type of decomposition, the concepts of benchmark and active risk are superimposed on the common factor and specific risks itemized in Total Risk Decomposition (*see* Figure 4-4).

- *Benchmark risk* is the risk associated with the benchmark portfolio.

- *Active risk* is the risk that arises from the manager's effort to outperform the benchmark. It is further divided into active specific and active common factor sources. Active risk is also known as tracking error.

This perspective is most commonly used in analyzing index funds as well as traditionally managed active portfolios.

In this type of decomposition, there is no concept of a market portfolio. The analysis concentrates solely on the managed portfolio against the benchmark that you select. However, for most managers, market risk is a component of active risk; these managers might prefer Active Systematic-Active Residual Risk decomposition, the fourth and last type.

Active Systematic-Active Residual Risk Decomposition

Finally, Active Systematic-Active Residual Risk decomposition (*see* Figure 4-5), the most complete perspective, expands the previous decomposition by including systematic sources of risk. Both this method and Active Risk are helpful in performance evaluation and analysis because they consider the *benchmark portfolio*, which reveals management style.

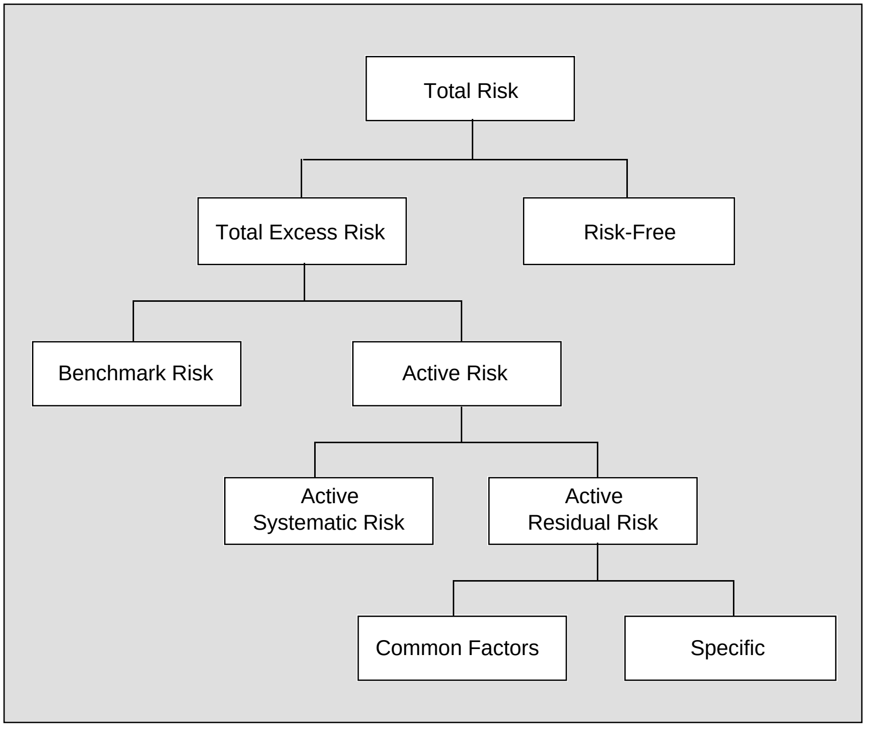


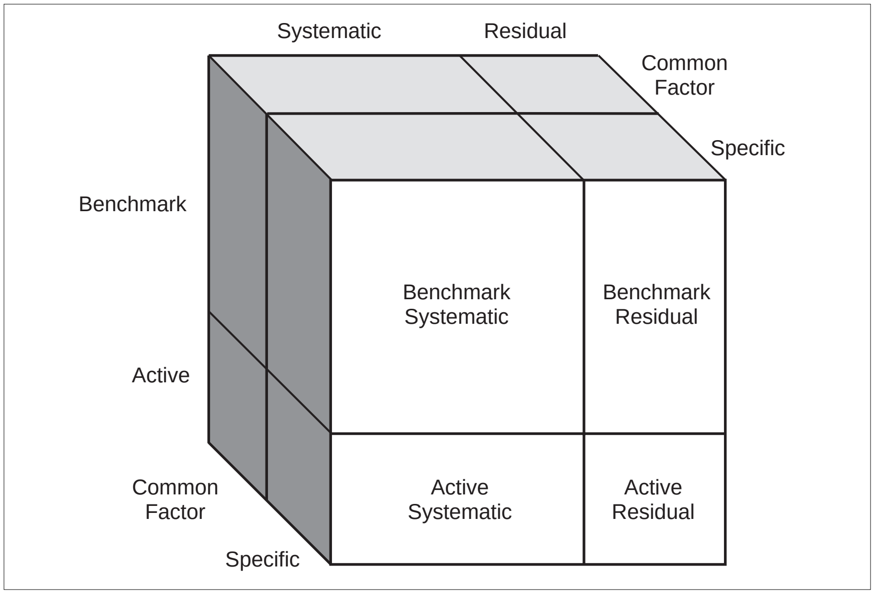
Figure 4-5
Active Systematic-Active Residual Risk Decomposition

This risk decomposition is useful for managers who overlay a market-timing strategy on their stock selection process and don't want market risk considerations to affect their analysis.

Summary of risk decomposition

These methods of risk decomposition represent compatible perspectives. Figure 4-6 shows how the four methods are different ways to slice the same pie—specific/common factor, systematic/residual, and benchmark/active.

Figure 4-6
Risk Decomposition Overview



Performance attribution

Performance attribution completes the portfolio management process by applying the MFM to past portfolio activity. Performance attribution is the process of matching return with its sources. Return is attributed to common factors, market timing, and asset selection. Using benchmark comparisons to judge performance, the value of each investment decision can be determined.

BARRA's performance attribution programs decompose return into its major components using any of the four methods of risk decomposition listed above. Performance can then be evaluated with respect to customized or industry-standard benchmark portfolios designed to compare managers with their own standards.

Summary

In this chapter we have outlined the general management approaches available to equity managers and discussed how BARRA MFMs can be utilized at various stages of the management process. In the next chapter we will describe in detail the process of building an MFM.

5. BARRA Multiple-Factor Modeling

A BARRA equity risk model is the product of a thorough and exacting model estimation process. This section provides a brief overview of model estimation.

Figure 5-1
Model Estimation Process

1. *Data acquisition and cleaning*
2. *Descriptor selection and testing*
3. *Descriptor standardization*
4. *Risk index formulation*
5. *Industry allocation*
6. *Factor return estimation*
7. *Covariance matrix calculation:*
   a. *Exponential weighting*
   b. *Market volatility: GARCH*
8. *Specific risk forecasting*
9. *Model updating*

Overview

The creation of a comprehensive equity risk model is an extensive, detailed process of determining the factors that describe asset returns. Model estimation involves a series of intricate steps that is summarized in Figure 5-1.

The first step in model estimation is acquiring and cleaning data. Both market information (such as price, dividend yield, or capitalization) and fundamental data (such as earnings, sales, or assets) are used. Special attention is paid to capital restructurings and other atypical events to provide for consistent cross-period comparisons.

Descriptor selection follows. This involves choosing and standardizing variables which best capture the risk characteristics of the assets. To determine which descriptors partition risk in the most effective and efficient way, the descriptors are tested for statistical significance. Useful descriptors often significantly explain cross-sectional returns.

Risk index formulation and assignment to securities is the fourth step. This process involves collecting descriptors into their most meaningful combinations. Though judgment plays a major role, a variety of techniques are used to evaluate different possibilities. For example, cluster analysis is one statistical tool used to assign descriptors to risk indices.

Along with risk index exposures, industry allocations are determined for each security. In most BARRA models a single industry exposure is assigned, but multiple exposures for conglomerates are calculated in a few models, including the U.S. and Japan models.

Next, through cross-sectional regressions, factor returns are calculated and used to estimate covariances between factors, generating the covariance matrix used to forecast risk. Exponential weighting of data observations may be used if testing indicates it improves accuracy. This matrix may be further modified to utilize GARCH techniques.

Specific returns are separated out at this stage and *specific risk* is forecast. This is the portion of total risk that is related solely to a particular stock and cannot be accounted for by common factors. The greater an asset's specific risk, the larger the proportion of return attributable to idiosyncratic, rather than common, factors.

Lastly, the model undergoes final testing and updating. Risk forecasts are tested against alternative models. Tests compare *ex ante* forecasts with *ex post* realizations of beta, specific risk, and active risk. New information from company fundamental reports and market data is incorporated, and the covariance matrix is recalculated.

Figure 5-2 summarizes these steps.

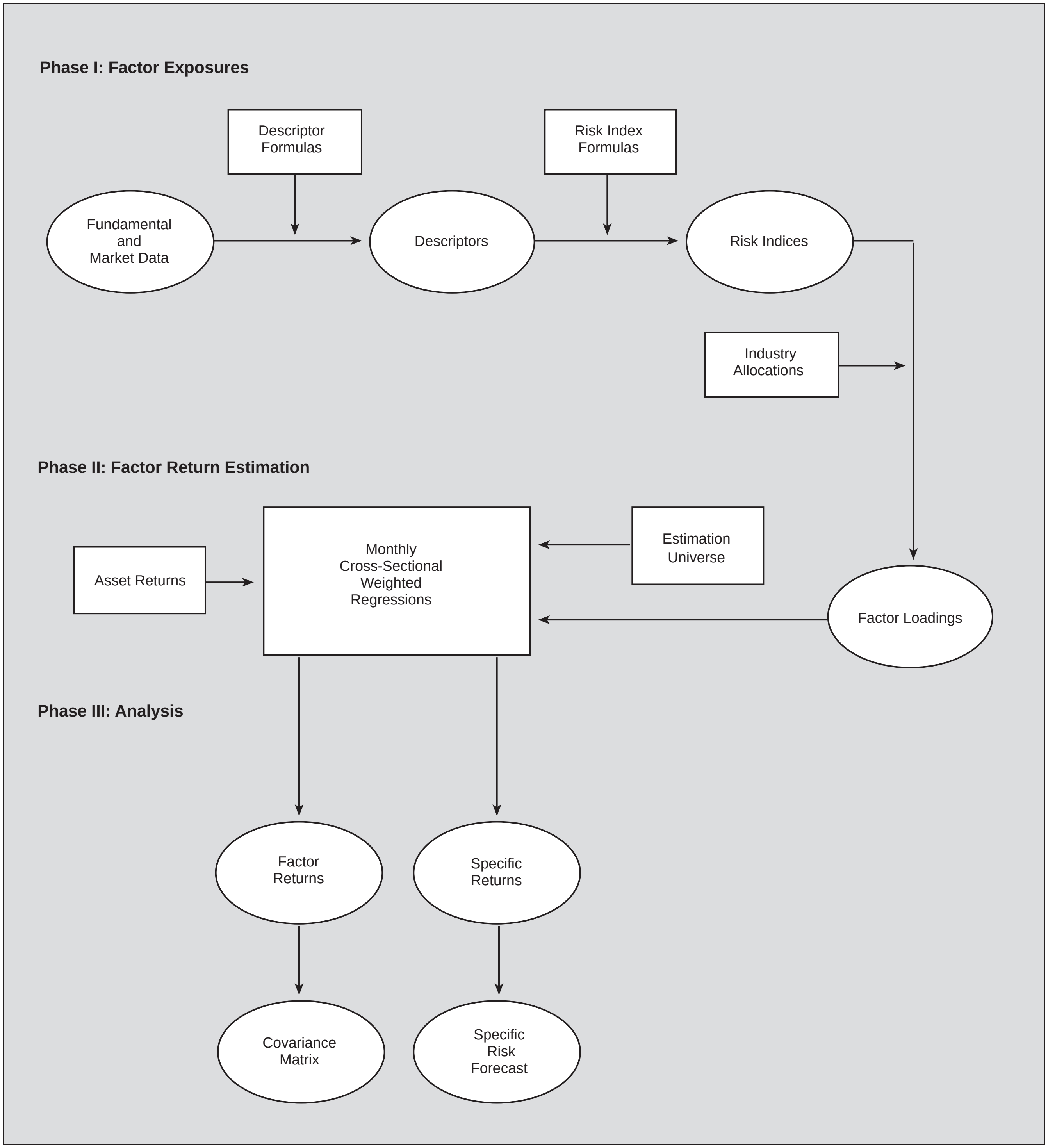


Figure 5-2

Data Flow for Model Estimation

This figure depicts the model estimation process. The oval shapes mark the data flow throughout model estimation, while the rectangular shapes show manipulations of and additions to the data.

Descriptor selection and testing

Descriptor candidates are drawn from several sources. Market information, such as trading volume, stock prices, and dividends, is available daily. Fundamental company data—such as earnings, assets, and industry information—are derived from quarterly and annual financial statements. For some descriptors, market and fundamental information is combined. An example is the earnings to price ratio, which measures the relationship between the market's valuation of a firm and that firm's earnings.

Descriptor selection is a largely qualitative process that is subjected to rigorous quantitative testing. First, preliminary descriptors are identified. Good descriptor candidates are individually meaningful; that is, they are based on generally accepted and well-understood asset attributes. Furthermore, they should divide the market into well-defined categories, providing full characterization of the portfolio's important risk features. BARRA has more than two decades of experience identifying important descriptors in equity markets worldwide. This experience informs every new model we build.

Selected descriptors must have a sound theoretical justification for inclusion in the model. They must be useful in predicting risk and based on timely, accurate, and available data. In other words, each descriptor must add value to the model. If the testing process shows that they do not add predictive power, they are rejected.

Normalization

Normalization is the process of setting random variables to a uniform scale. Also called *standardization*, it is the process by which a constant (usually the mean) is subtracted from each number to shift all numbers uniformly. Then each number is divided by another constant (usually the standard deviation) to shift the variance.

Descriptor standardization

The risk indices are composed of descriptors designed to capture all the relevant risk characteristics of a company. The descriptors are first normalized; that is, they are standardized with respect to the estimation universe. This is done to allow combination of descriptors into meaningful risk factors, known as *risk indices*. The normalization process is summarized by the following relation:

$$\left[\textit{normalized descriptor}\right] = \frac{\left[\textit{raw descriptor}\right] - \left[\textit{mean}\right]}{\left[\textit{standard deviation}\right]}$$

Risk index formulation

We regress asset returns against industries and descriptors, one descriptor at a time, after the normalization step. Each descriptor is tested for statistical significance. Based on the results of these calculations and tests, descriptors for the model are selected and assigned to risk indices.

Risk index formulation is an iterative process. After the most significant descriptors are added to the model, remaining descriptors are subjected to stricter testing. At each stage of model estimation, a new descriptor is added only if it adds explanatory power to the model beyond that of industry factors and already-assigned descriptors.

Industry allocation

For most BARRA equity models, companies are allocated to single industries. For the U.S. and Japan, however, sufficient data exist to allocate to multiple industries.

For the U.S. and Japan, industry exposures are allocated using industry segment data (i.e., operating earnings, assets, and sales). The model incorporates the relative importance of each variable in different industries. For example, the most important variable for oil companies would be assets; for retail store chains, sales; and for stable manufacturing companies, earnings. For any given multi-industry allocation, the weights will add up to 100%.

Multiple industry allocation provides more accurate risk prediction and better describes market conditions and company activity. BARRA's multiple-industry model captures changes in a company's risk profile as soon as new business activity is reported to shareholders. Alternative approaches can require 60 months or more of data to recognize changes that result from market prices.

Factor return estimation

The previous steps have defined the exposures of each asset to the factors at the beginning of every period in the estimation window. The factor excess returns over the period are then obtained via a cross-sectional regression of asset excess returns on their associated factor exposures:

$$\tilde{r}_t = X_t \tilde{f}_t + u_t \qquad \text{(EQ 5-1)}$$

where:

$\tilde{r}_t$ = excess returns to each asset

X_t = exposure matrix of assets to factors

$\tilde{f}_t$ = factor returns to be estimated

u_t = specific returns

The resulting factor returns are robust estimates which can be used to calculate a factor covariance matrix to be used in the remaining model estimation steps.

Covariance matrix calculation

The simplest way to estimate the factor covariance matrix is to compute the sample covariances among the entire set of estimated factor returns. Implicit in this process is the assumption that we are modeling a stable process and, therefore, each point in time contains equally relevant information.

There is evidence, however, that correlations among factor returns change. Moreover, a stable process implies a stable variance for a well-diversified portfolio with relatively stable exposures to the factors. There is considerable evidence that, in some markets, the volatility of market index portfolios changes. For example, periods of high volatility are often followed by periods of high volatility. The changing correlations among factor returns, and the changing volatility of market portfolios, belie the stability assumption underlying a simple covariance matrix.

For certain models we relax the assumption of stability in two ways (*see* Table 5-1 at the end of this chapter for details). First, in computing the covariance among the factor returns, we assign more weight to recent observations relative to observations in the distant past. Second, we estimate a model for the volatility of a market index portfolio—for example, the S&P 500 in the U.S. and the TSE1 in Japan—and scale the factor covariance matrix so that it produces the same volatility forecast for the market portfolio as the model of market volatility.

Exponential weighting

Suppose that we think that observations that occurred 60 months ago should receive half the weight of the current observation. Denote by T the current period, and by t any period in the past, $t = 1,2,3,\ldots,T\text{-}1,T$, and let $\delta = .5^{1/60}$. If we assign a weight of δ^{T-t} to observation t, then an observation that occurred 60 months ago would get half the weight of the current observation, and one that occurred 120 months ago would get one-quarter the weight of the current observation. Thus, our weighting scheme would give *exponentially declining weights* to observations as they recede in the past.

Our choice of sixty months was arbitrary in the above example. More generally, we give an observation that is *HALF-LIFE* months ago one-half the weight of the current observation. Then we let:

$$\delta = (.5)^{\frac{1}{HALFLIFE}} \qquad \text{(EQ 5-2)}$$

and assign a weight of:

$$w(t) = \delta^{T-t}. \qquad \text{(EQ 5-3)}$$

The length of the *HALF-LIFE* controls how quickly the factor covariance matrix responds to recent changes in the market relationships between factors. Equal weighting of all observations corresponds to *HALF-LIFE* = ∞. Too short a *HALF-LIFE* effectively "throws away" data at the beginning of the series. If the process is perfectly stable, this decreases the precision of the estimates. Our tests show that the best choice of *HALF-LIFE* varies from country to country. Hence, we use different values of *HALF-LIFE* for different single country models.

Computing market volatility: Extended GARCH models

There is considerable evidence that, in some markets, market volatility changes in a predictable manner. We find that returns that are large in absolute value cluster in time, or that volatility persists. Moreover, periods of above-normal returns are, on average, followed by lower volatility, relative to periods of below-normal returns. Finally, we find that actual asset return distributions exhibit a higher likelihood of extreme outcomes than is predicted by a normal distribution with a constant volatility.

Variants of GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models capture these empirical regularities by allowing volatility to increase following periods of high realized volatility, or below-normal returns, and allowing volatility to decrease following periods of low realized volatility, or above-normal returns.

The following discussion lays out the general theory of extended GARCH modeling. Variants of this approach will be applied as appropriate to BARRA single country models over time. *See* Table 5-1 of this chapter for current coverage.

Formally, denote by $\tilde{r}_t$ the market return at time *t*, and decompose it into its expected component, $\mathrm{E}(\tilde{r}_t)$, and a surprise, ε_t:

$$\tilde{r}_t = \mathrm{E}(\tilde{r}_t) + \varepsilon_t \qquad \text{(EQ 5-4)}$$

The observed persistence in realized volatility indicates that the variance of the market return at *t*, $Var(\tilde{r}_m)_t$, can be modeled as:

$$Var(\tilde{r}_m)_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta\, Var(\tilde{r}_m)_{t-1} \qquad \text{(EQ 5-5)}$$

This equation, which is referred to as a GARCH(1,1) model, says that current market volatility depends on recent realized volatility via ε_{t-1}^2, and on recent forecasts of volatility via $Var(\tilde{r}_m)_{t-1}$. If α and β are positive, then this period's volatility increases with recent realized and forecast volatility.

GARCH(1,1) models have been found to fit many financial time series. Nevertheless, they fail to capture relatively higher volatility following periods of below-normal returns. We can readily extend the GARCH(1,1) model to remedy this shortcoming by modeling market volatility as:

$$Var(\tilde{r}_m)_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta\, Var(\tilde{r}_m)_{t-1} + \theta\varepsilon_{t-1} \qquad \text{(EQ 5-6)}$$

If θ is negative, then returns that are larger than expected are followed by periods of lower volatility, whereas returns that are smaller than expected are followed by higher volatility.

Having satisfactorily fit a GARCH model to the volatility of a market proxy portfolio, it is used to scale the factor covariance matrix so that the matrix gives the same risk forecast for the market portfolio as the GARCH model. In implementing the scaling, however, we scale only the systematic part of the factor covariance matrix.

Specific risk modeling

Overview

Referring to the basic factor model:

$$\tilde{r}_i = \sum_k X_{ik} \tilde{f}_k + \tilde{u}_i \qquad \text{(EQ 5-7)}$$

The specific risk of asset *i* is the standard deviation of its specific return, $Std(\tilde{u}_i)$. The simplest way to estimate the specific risk matrix is to compute the historical variances of the specific returns. This, however, assumes that the specific return distribution is stable over time. Rather than use historical estimates, we *model* specific risk to capture fluctuations in the general level of specific risk and the relationship between specific risk and asset fundamental characteristics.

An asset's specific risk is the product of the *average* level of specific risk that month across assets, and each asset's specific risk *relative* to the average level of specific risk. Moreover, our research has shown that the relative specific risk of an asset is related to the asset's fundamentals. Thus, developing an accurate specific risk model involves a model of the average level of specific risk across assets, and a model that relates each asset's relative specific risk to the asset's fundamental characteristics.

Methodology

Denote by S_t the average level of specific risk across assets at time t, and by V_{it} asset i's specific risk relative to the average.

In equation form:

$$Std\left(\tilde{u}_{it}\right) = S_t\left(1 + V_{it}\right) \qquad \text{(EQ 5-8)}$$

where:

$Std\left(\tilde{u}_{it}\right)$ = asset specific risk,

S_t = average level of specific risk at time t, and

V_{it} = asset i's specific risk relative to the average.

We estimate a model for S_t via time-series analysis, in which the average level of realized specific risk is related to its lagged values and to lagged market returns. Similarly, we estimate a model of relative specific risk by regressing realized relative specific risks of all firms, across all periods, on the firm fundamentals, which include the BARRA risk factors.

Modeling the average level of specific risk

We model the average level of specific risk via a time series model, where the average specific risk is related to its own lagged values, as well as to lagged market excess returns:

$$S_t = \alpha + \sum_{i=1,k} \beta_i S_{t-i} + \beta_{k+1} r_{m_{t-1}} + \varepsilon_t \qquad \text{(EQ 5-9)}$$

where:

S_t = average specific risk at time t,

α = expected level of average specific risk across assets,

β_i = sensitivity of average specific risk to lagged values,

S_{t-i} = lagged average specific risk,

β_{k+1} = sensitivity of average specific risk to market returns,

$r_{m_{t-1}}$ = market return at time t-1, and

ε_t = residual level of specific risk.

This simple model captures mean reversion or trends and persistence in volatility, as well as lower average specific risk following market rises, and higher specific risk following market declines (provided β_{k+1} is negative). We evaluate the performance of alternative weighting schemes by regressing realized average specific risk against predicted specific risk out of sample:

$$S_t = a + b\hat{S}_t + e_t \tag{EQ 5-10}$$

The results of these tests help us determine the appropriate coefficients in Equation 5-9.

Modeling the relative level of specific risk

To model the relative level of specific risk, we first identify factors that may account for the variation in specific risk among assets. These factors will vary depending on the country model and may include:

- Industry membership
- Risk index exposures

Having identified these factors, we then estimate the relationship by performing the "pooled" *cross-sectional* regression:

$$V_{it} = \sum_{k=1,K} X_{ikt}\gamma_k + e_{it} \tag{EQ 5-11}$$

where:

V_{it} = relative specific risk for asset i at time t,

t = 1 to T months,

i = 1 to N assets,

X_{ikt} = the exposure of asset i to factor k at time t, and

γ_k = factor k's contribution to relative specific risk.

In estimating this relationship, we mitigate the effects of outliers on the estimators by Windsorizing descriptors and risk indices.

Estimating the scaling coefficients

Average specific risk can vary widely over the full capitalization range of an equity market. To correct for this effect, scaling coefficients are used to adjust for any specific risk bias. To estimate the scaling coefficients, we divide the set of securities into capitalization deciles and, for each decile, we compute the bias in predicted specific risk from the previous steps. The scaling coefficients are then computed as a piece-wise linear function of an asset's relative capitalization in a manner that makes the average bias in predicted specific risk zero for each capitalization decile.

Final specific risk forecast

The above three components—average level of specific risk, relative level of specific risk, and decile scaling coefficients—are combined to produce the final asset specific risk forecast:

$$Std(\varepsilon_{it}) = scale_{it} \cdot \hat{S}_t \cdot (1 + \hat{V}_{it}) \qquad \text{(EQ 5-12)}$$

where:

ε_{it} = specific risk of asset ***i*** at time ***t***,

$scale_{it}$ = scale coefficient for the decile that asset ***i*** falls into at time ***t***,

$\hat{S}_t$ = average level of specific risk across all assets, and

$\hat{V}_{it}$ = relative level of specific risk for asset ***i*** at time ***t***.

Updating the model

Model updating is a process whereby the most recent fundamental and market data are used to calculate individual stock exposures to the factors, to estimate the latest month's factor returns, and to recompute the covariance matrix.

The latest data are collected and cleaned. Descriptor values for each company in the database are computed, along with risk index exposures and industry allocations. Next, a cross-sectional regression is run on the asset returns for the previous month. This generates factor returns which are used to update the covariance matrix. Finally, this updated information is distributed to users of BARRA's applications software.

Comparison of risk model features

Table 5-1 summarizes BARRA's single country equity models and the features of each as of January 1998.

Country Model	Number of Industries	Number of Risk Indices	Industry Allocation Method: Multiple/Single	Exponential Smoothing Half-Life: Number of Months	GARCH: Yes/No
Australia (AUE2)	24	9	Single	90	No
Canada (CNE3)	40	11	Single	60	No
France (FRE3)	12	9	Single	48	No
Germany (GRE2)	17	10	Single	90	No
Germany—Trading Model (GRTM)	17	10	Single	90	Yes
Hong Kong (HKE1)	13	10	Single	48	No
Japan (JPE2)	40	12	Multiple	60	Yes
Korea (KRE1)	28	12	Single	24	No
Malaysia (MLE1)	14	10	Single	36	No
Netherlands (NLE1)	8	7	Single	60	No
New Zealand (NZE1)	6	5	Single	48	Yes
South Africa (SAE1)	23	11	Single	36	No
Sweden (SNE3)	20	10	Single	90	Yes
Switzerland (SWE2)	12	8	Single	90	No
Taiwan (TWE1)	25	10	Single	36	No
Thailand (THE1)	32	9	Single	36	No
U.K. (UKE5)	38	12	Single	90	No
U.K.—Trading Model (UKTM)	38	12	Single	90	Yes
U.S. (USE2)	55	13	Multiple	90	Yes
U.S. (USE3)	52	13	Multiple	90	Yes
U.S.—Small Cap (USSC)	42	11	Single	90	No
U.S.—Short-Term Risk (STRM)	55	3	Multiple	40 trading days	Yes

Table 5-1 BARRA Single Country Equity Risk Models* †

* As of 1/98.
† See text for definitions of model features.

6. Advantages of US-E3 Over US-E2

Overview

The original BARRA product was a U.S. equity risk model, developed in 1975. Dubbed US-E1, it was superseded by a second-generation model, US-E2, in the early 1980s. US-E2 has served BARRA clients well over the last 18 years, but we identified a variety of deficiencies which made a significant upgrade desirable. The improvements that have been implemented in the U.S. Equity Model Version 3 (US-E3) cumulatively result in a demonstrably better model, both in terms of qualitative features and quantitative measures.

Highlights of US-E3's advances over US-E2:

- complete industry reclassification
- flexible industries
- larger estimation universe
- addition of Size Nonlinearity risk index
- improved independence between risk indices
- improved specific risk model
- improved GARCH model
- calculation of REIT risk index exposures
- higher *t*-stats in general
- improved data quality assurance
- improved model diagnostics
- annual re-estimation of specific risk, GARCH, and descriptor weight parameters

Industries

Reclassification

Experience from US-E2, client input, and the changing nature of the market led to a complete reconsideration of BARRA's industry classification. Once we decided upon a revised set of industries, we performed a company-by-company review of available tear sheets for all assets in the US-E2 estimation universe (the HICAP) by at least two people. Even if US-E2 and US-E3 industries mapped one-to-one or many-to-one, assets may have moved to other US-E3 industries if tear sheet review indicated they had been misclassified within US-E2.

Here are some highlighted differences between US-E2 and US-E3 classification schemes:

- US-E3 does not have a Miscellaneous industry, as US-E2 does. All assets must now be assigned to specific industries.
- US-E2's five oil-related industries have been collapsed into three industries for US-E3, to reflect the massive changes in this segment since the early 1980s.
- US-E2's Services was a catchall industry. US-E3 has Information Services (more high-tech services) and Industrial Services (more low-tech types of industrial services). Also, software companies lumped into US-E2 Services are now in their own industry, Computer Software.
- US-E2's Aluminum, Coal & Uranium, and Iron & Steel have been combined into US-E3's Mining & Metals.
- US-E2's Forest Products and Paper are a single industry for US-E3.
- Photography is not a US-E3 industry.
- Containers is not a US-E3 industry.
- US-E3 has added the industries Entertainment, Semiconductors, Computer Software, Wireless Telecommunications, and Securities & Asset Management.
- US-E3 industries distinguish between Medical Providers & Services and Medical Products.

Flexible industries

This is a major advance in US-E3. BARRA risk models have historically had great difficulty adding or subtracting industries. US-E3 assigns industries to assets based on 82 mini-industries. This mini-industry classification has been performed historically for each asset from January 1975 to the present. This allows us to watch for an increase in assets in a given mini-industry. Once a minimum threshold has been reached, the mini-industry can be promoted to industry status if it appears appropriate. By "appropriate" we mean that either (a) there are statistical reasons for estimating the mini-industry separately; and/or (b) there is market/client differentiation which makes the mini-industry ripe for industry classification. Some element of judgment will be exercised here in consultation with our clients.

We will also be monitoring the decline of assets within existing industries. Too few for too long will trigger an extinction of the industry. A situation similar to Coal & Uranium (which currently has no assets in the HICAP with this as primary industry) will not occur in US-E3.

US-E3 performance analysis (as implemented in BARRA's Windows-based Aegis System) will be able to handle these rising and falling industries. Each industry will be assigned to one of 13 permanent economic sectors which will maintain continuity for all assets even as their industries change.

Increased size of the estimation universe

US-E3 has increased the size of its estimation universe by roughly 50%. This means that more of the midsize companies (those approximately 1,001–2,000 in capitalization rank) are participating in calculation of the factor returns. Since a greater number of BARRA clients are investing in such companies, this is a desirable feature.

We extended the estimation universe only after we determined that (a) coverage of the descriptors used for risk index calculation would not deteriorate and (b) the factor returns would not change "too much" by the addition of lower-capitalization assets.

Risk indices

In general, US-E3 has pared down the total number of descriptors used for risk index construction. This has been done in order to reduce the number of descriptors used within each risk index so that only the most significant data are used and to remove overlaps which occurred in US-E2.

Size Nonlinearity factor

This is the most novel addition to the risk indices. Internal research has shown that US-E2, while doing a good job of risk forecasting of the top 500–750 capitalization assets, was systematically underpredicting risk for smaller capitalization assets. In other words, size has a nonlinear risk component. With the increase of the estimation universe to roughly the top 1,500 capitalization assets, this becomes even more of an issue. US-E3 uses a polynomial model to approximate this nonlinearity within the context of a linear model. The result is superior risk forecasts for assets and portfolios over the entire range of size-sorted portfolios.

Leverage

US-E3 calculates the Leverage exposure in the same fashion across all industries. US-E2 used (a) interest rate sensitivity for financial, utility, and railroad industries and (b) fundamental-based descriptors for all other industries. It is not a good idea to combine two different calculations under the same umbrella; for instance, there is no non-arbitrary way to normalize them. We also found that there was no need to treat this index differently for different industries; it did not capture any added nuances.

Simpler volatility calculation

US-E2 uses a three-tiered method of computing the Variability in Markets risk index (which is the counterpart of US-E3's Volatility). Different descriptors and weights were used within each of US-E2's tiers. The tiers are based on: assets with options, HICAP assets without options, and all other assets. Historically, assets have jumped tiers, which resulted in noticeable and non-intuitive changes in Variability in Markets exposures from month to month.

US-E3 will not duplicate the three-tiered method of US-E2. Research during the development of US-E3 found no evidence to indicate that this "tiering" is necessary for providing accurate risk forecasts. Using a single set of core descriptors in addition to Option-Implied Standard Deviation (where appropriate), US-E3 achieves higher explanatory power than the US-E2 methodology.

Elimination of US-E2's Labor Intensity and Foreign Income

Although the motivation behind these risk indices is intuitive, coverage of data necessary for these risk indices is decidedly lower than for other risk indices. Since no substitute could be identified for Labor Intensity, we eliminated it.

US-E3's Currency Sensitivity differs from US-E2's Foreign Income through the use of a regression technique to determine directly the effect of currency fluctuations on asset returns. This implied sensitivity to foreign operations avoids the need to find fundamental data describing this effect and was found in our testing to be robust.

Improved independence between risk indices

In US-E2 several risk indices shared the same descriptors. This resulted in a certain amount of unnecessary multicollinearity—most notably related to Growth, Yield, and Earnings/Price. US-E3 risk indices consist of separate sets of descriptors for each risk index. The model benefits from this through more precise factor returns estimates (smaller standard errors, higher *t*-stats) which results in more precise common factor covariance matrix estimation.

Risk forecasting

There are two sources of US-E3's improvement in risk forecasting:

Improved GARCH model

The extended GARCH(1,1) model provides for a more accurate risk forecast of overall market volatility. It responds appropriately to observed asymmetry in market volatility related to up market moves and down market moves. For additional details, *see* Chapter 5. BARRA Multiple-Factor Modeling.

Specific risk model

US-E3 builds its specific risk model based on two risk forecasts: market average specific risk and relative specific risk. US-E2 builds its model based only on relative specific risk. This will make US-E3's specific risk predictions more accurate. *See* Chapter 5. BARRA Multiple-Factor Modeling for details.

Model fit-related issues

Factor return estimation

Even though the R-squared for the monthly factor return regressions are virtually the same for US-E2 and US-E3, the new industry classification and risk index modeling has resulted in more precise factor return estimates (which are reflected in increased *t*-statistics for the estimated factor returns). More precise factor return estimates provide better data for common factor risk forecasts.

Ongoing diagnostics

US-E3 will collect and save the set of information used in model-building each month as the model moves forward in time. This means BARRA can reassess modeling issues by simply examining this database, rather than having to redo historic estimations in the future. This makes it feasible to re-estimate model-related parameters (such as descriptor weights within risk indices and GARCH parameters) much faster.

Scheduled refitting of model parameters

US-E2 has always used the same weights for descriptors. The same is true for the parameters of the specific risk model. The US-E2 GARCH parameters have never been re-estimated. US-E3 will refit these models on a yearly or biyearly schedule. This will assure that the model responds to changes in the market. It will also provide an ongoing assessment of our modeling technique's true (live mode) out-of-sample performance.

Asset class issues

REITs

Most of the REITs in US-E2 have "0" for all of their risk index exposures except LOCAP. US-E3 calculates risk index exposures for its REITs. US-E3 also does a better job of separating the property REITs from the mortgage REITs.

Coming soon

After the release of US-E3 we will conduct extensive research into the modeling of mutual funds, bonds, preferred stock, SPIDERs, ADRs, IPOs, and other asset classes.

7. The US-E3 Estimation Universe

Overview

The primary attribute determining an asset's inclusion in the US-E3 estimation universe is its capitalization. The top 1,500 assets by capitalization comprise the heart of the monthly estimation universe. Additional assets are added to ensure depth within industries. This has resulted in a final universe size of around 1,900, which will fluctuate from month to month. Only U.S. stocks with Compustat data are considered. Any asset which drops below the capitalization cutoff will be given a grace period of 18 months before it is dropped from the estimation universe.

Selection process

There are six considerations involved:

1. S&P 500 membership
2. Compustat data present
3. Capitalization
4. Industry fill-in
5. Price
6. Grandfathering

The above considerations apply only to "plain vanilla" U.S. stocks. No REITs, ADRs, closed-end funds, etc. will be considered for inclusion in the estimation universe. The only possible exception here is in the case of S&P 500 membership. We will include any type of asset that is in the S&P 500, as long as we have assigned an industry to it. Considering only plain stocks for estimation universe membership ensures having the necessary industry assignments.

1. S&P 500 membership

All assets in the S&P 500 will be included in the estimation universe, even if other rules discussed below are violated. The capitalization and price considerations that follow do not apply to these assets.

2. Compustat data present

Since so many of the descriptors we need come from Compustat data, we do not want to select assets for our estimation universe when we know we will have to resort to replacement rules for all of the descriptors for those assets. Accordingly, we will use only assets that have annual Compustat data within at most 21 months from the model update.

3. Capitalization

We will automatically include the top 1,500 stocks. No ADRs will be included *unless* they are part of the S&P 500.

4. Minimum price

This will be applied only to assets that are selected for industry fill-in (described in the next step). No asset less than $5.00 per share will be included unless it falls under 1. or 3. above.

5. Industry fill-in

There should be at least 20 assets in each industry, if possible. After collecting the top 1,500 assets, we will determine which industries are thin. Each industry will be filled in, searching from the highest remaining capitalization assets in descending order, until there are 20 assets within the thin industries. It may not be possible to obtain all the assets we want. To avoid including companies that are too small, we will exclude from consideration any asset in the lowest 10% of capitalization.

6. Grandfathering

Assets that are near the top 1,500 capitalization cutoff bounce in and out of the estimation universe from month to month. If this were allowed, their risk forecasts would bounce as well, since their Non-Estimation Universe (NONESTU) exposure would be changing. We will minimize this by allowing an 18-month grace period. If an asset was in the estimation universe by virtue of either 1., 3., or 5. above and drops out due to falling market capitalization, the asset will drop out of the estimation universe after 18 months unless it recovers to the cutoff during that time. This will minimize the associated jumps in risk forecasts.

Even grandfathered assets must pass the minimum price condition and bottom 10th percentile in capitalization condition.

This methodology *requires* keeping a time series database of assets that were part of the estimation universe based on contemporaneous considerations.

Comparison with the US-E2 estimation universe

In general, most of the US-E2 HICAP is included in US-E3's estimation universe. Differences between the two universes arise principally from the following:

1. US-E3 takes the top 1,500 capitalization assets as its starting point. In US-E2, the HICAP takes the top 1,000 assets as its starting point. US-E3, therefore, is letting a larger number of assets participate in the estimation of industry and risk index factor returns. Studies addressing the effect of expanding the size of the estimation universe with respect to deterioration of coverage of descriptors and the resulting factor returns led to the selection of the number 1,500.

2. The grandfathering and the monthly quest for industry depth have not been done in a consistent manner for US-E2. By following a rules-based process as described above, we will improve consistency.

3. US-E2 has some very small assets in the HICAP. The bottom tenth percentile rule for US-E3 prevents this.

8. US-E3 Risk Indices and Descriptors

Differences between US-E2 and US-E3 risk indices

General differences

In US-E2, there are a number of instances of the same descriptor appearing in more than one risk index. For example, descriptors relating earnings to price comprise part of the Growth risk index as well as the estimated Earnings-to-Price Ratio risk index, share turnover ratios appear in the Trading Activity risk index as well as the Variability in Markets risk index for some companies, and Dividend Yield is a separate risk index as well as a part of the Growth risk index. In some cases—for example, Yield and Growth—this leads to correlated exposures resulting in lower precision for the estimates of factor returns corresponding to these factors. To avoid highly correlated risk index exposures in US-E3, we decided to assign each descriptor to a single risk index. For most descriptors, the risk indices that they belonged to was fairly obvious; for the not-so-obvious cases, we assigned descriptors to risk indices based on a study we conducted.

Specific differences

Specific differences between US-E2 and US-E3 risk indices are as follows:

Volatility

In US-E2, this risk index has a three-tiered structure, with different ways of constructing this risk index for optioned stocks, other exchange-traded stocks, and thinly traded stocks. In US-E3, this risk index is constructed using a single rule for all stocks. For stocks that have options, option-implied standard deviation is computed using the Black-Scholes formula. For stocks without options, the option-implied standard deviation descriptor is treated as missing, and a replacement rule is used to estimate the descriptor value.

Size

In US-E2, the Size risk index is constructed using a weighted sum of the log of market capitalization and the log of total assets of the company. In contrast, US-E3 measures size exposure using simply the log of market capitalization. This measure of size exposure is consistent with the well-documented “size effect,” the term used by academics to refer to the strong linear relationship between average returns and log of market capitalization over many time periods.

Size Nonlinearity (US-E3)

This is a new risk index in US-E3 and does not appear in US-E2. The estimation universe in US-E3 consists of the top 1,500 stocks ranked by capitalization plus smaller companies chosen to ensure a reasonable number of companies in each industry. Our research showed that the linear relationship between the cross-section of returns and log of market capitalization is a good approximation for larger companies (especially in the top 500-750 range). For smaller companies, our research showed that there are some non-linearities in the relation between returns and log of market capitalization. The Size Nonlinearity risk index ensures that the US-E3 model does a good job of capturing the relationship between returns and size for all companies, not just the larger ones.

Growth

In US-E2, this risk index is a combination of historical earnings and asset growth, historical payout ratios, analyst-predicted growth, earnings- to-price, and dividend yield variables. The weights of each descriptor were based on a predictive model for five-year-ahead earnings growth. In US-E3, this risk index is a combination of historical earnings and asset growth, historical payout ratios, and analyst-predicted earnings growth. Earnings yield and dividend yield variables do not appear in the Growth risk index in US-E3. The weights of each descriptor are based on a predictive model for three-year-ahead sales growth. We estimated other predictive models for five-year-ahead sales growth and five-year-ahead earnings growth. The three-year sales growth model that we estimated provided the most intuitive set of weights, had the best in-sample fit, had good out-of-sample properties, and explained the cross-section of returns well.

Financial Leverage (US-E2) and Leverage (US-E3)

In US-E2, this risk index is constructed differently for companies in finance-related industries, railroads, and utilities. For companies in these industries, a market-based interest-rate-sensitivity measure is used to measure the exposure to the leverage factor. In contrast, US-E3 uses the same rule, based on the usual measures of book leverage, market leverage, etc., for all companies. Intuition suggests that the usual measures of leverage for companies in finance-related industries and utilities are likely to be very high, leading to clustering of such companies at the high end of the leverage spectrum. However, this intuition is not borne out by inspection of the data. In most cases, our research found that these companies have leverage values comparable with other companies, so the usual measures of leverage can be used as measures of exposure to the Leverage factor.

Foreign Income (US-E2) and Currency Sensitivity (US-E3)

In US-E2, this risk index is constructed as the proportion of operating income of a company that arises from non-domestic sources. Many companies do not report their operating income by geographic regions, so US-E2 uses replacement rules to fill in values for companies that do not report this information. Data coverage for this information has been a persistent problem for US-E2. In addition, US-E2 did not account for any currency hedging. To avoid these problems, US-E3 has adopted a different measure of non-domestic risk based on currency exchange rate returns. This measure is a regression-based sensitivity of asset returns to exchange rate returns, and was shown in our tests to be robust.

Labor Intensity (US-E2)

In US-E2, this risk index uses descriptors based on labor expenses, total assets, net plant and gross plant of a company. A large number of companies do not report labor expenses as a separate item in their financial statements. This leads to the use of a replacement rule for the descriptor based on labor expenses. A related issue is that over the history of the US-E2 model, the Labor Intensity risk index appears to be only marginally significant. We therefore decided to drop the Labor Intensity risk index in US-E3.

US-E3 and US-E2 risk indices at a glance

The table below displays US-E3 and US-E2 risk index information in a concise form. Each US-E3 risk index is paired with its most-closely-associated counterpart in US-E2, and a list of descriptors that make up each risk index is provided. US-E2 descriptors that appear with a strikethrough (for example, ~~log of total assets~~) are those that are dropped from the corresponding US-E3 risk index. Such descriptors may still be a part of some other risk index in US-E3 or may have been dropped from the model. US-E3 descriptors that appear in a shaded box are new descriptors not present in US-E2.

Descriptors in US-E3 Risk Index	US-E3 Risk Index	US-E2 Risk Index	Descriptors in US-E2 Risk Index
Beta times sigma	Volatility	Variability in Markets	Beta times sigma
		Note: This is a three-tiered risk index in US-E2. The descriptors shown at right appear in at least one of the "tiers."	
Daily standard deviation			Daily standard deviation
High-low price			Log of stock price
Log of stock price			Cumulative range
Cumulative range			Serial dependence
Volume beta			Option-implied standard deviation
Serial dependence			~~Volume to variance~~
Option-implied standard deviation			~~Share turnover rate (annual)~~
Relative strength	Momentum	Success	Relative strength
Historical alpha			Historical alpha
			~~Dividend cuts over the last five years~~
			~~Recent earnings change~~
			~~Analyst-predicted earnings growth~~
			~~Growth in earnings per share~~
Log of market capitalization	Size	Size	Log of market capitalization
			~~Log of total assets~~
			~~Indicator of earnings history~~
Cube of log of market capitalization	Size Nonlinearity		

Table 8-1

US-E3 and US-E2 Risk Indices

Descriptors in US-E3 Risk Index	US-E3 Risk Index	US-E2 Risk Index	Descriptors in US-E2 Risk Index
Share turnover rate (annual)	Trading Activity	Trading Activity	Share turnover rate (annual)
Share turnover rate (quarterly)			Share turnover rate (quarterly)
Share turnover rate (monthly)			Share turnover rate (five years)
Share turnover rate (five years)			~~Number of analysts~~
Indicator for forward split			Volume to variance
Volume to variance			
Payout ratio over five years	Growth	Growth	Payout ratio over five years
Variability in capital structure			Variability in capital structure
Growth rate in total assets			Growth rate in total assets
Earnings growth rate over the last five years			Earnings growth rate over the last five years
Analyst-predicted earnings growth			Analyst-predicted earnings growth
Recent earnings change			Recent earnings change
			~~Earnings-to-price ratio (five years)~~
			~~Normalized earnings-to-price ratio~~
			~~Dividend yield (five years)~~
			~~Yield forecast~~
			~~Indicator of zero yield~~
			~~Earnings to price ratio~~
			~~Analyst-predicted earnings-to-price ratio~~
Analyst-predicted earnings-to-price	Earnings Yield	Earnings-to-Price Ratio	Analyst-predicted earnings-to-price

Table 8-1
US-E3 and US-E2 Risk Indices

Descriptors in US-E3 Risk Index	US-E3 Risk Index	US-E2 Risk Index	Descriptors in US-E2 Risk Index
Trailing annual earnings-to-price Historical earnings-to-price			Earnings-to-price Historical earnings-to-price
Book-to-price ratio	Value	Book-to-Price Ratio	Book-to-price ratio
Variability in earnings Variability in cash flows Extraordinary items in earnings Standard deviation of analyst-predicted earnings-to-price	Earnings Variability	Earnings Variability	Variability in earnings Variability in cash flows Extraordinary items in earnings Standard deviation of analyst-predicted earnings-to-price ~~Earnings covariability~~ ~~Concentration~~
Market leverage Book leverage Debt to total assets Senior debt rating	Leverage	Financial Leverage *Note:* For finance-related industries, utilities, and railroads, this risk index is constructed using a measure of bond market sensitivity.	Book leverage Debt to total assets ~~Uncovered fixed charges~~ ~~Bond market sensitivity~~
Exposure to foreign currencies	Currency Sensitivity		
		~~Foreign Income~~	~~Proportion of operating income from foreign sources~~
		~~Labor Income~~	~~Labor share~~ ~~Inflation-adjusted plant to equity~~ ~~Net plant to gross plant~~
Predicted dividend yield	Dividend Yield	Yield	Predicted dividend yield
Indicator for firms outside US-E3 estimation universe	Non-Estimation Universe Indicator	LOCAP Indicator	Indicator for firms outside US-E2 HICAP universe

Table 8-1

US-E3 and US-E2 Risk Indices

Risk index definitions

1. Volatility

This risk index captures relative volatility using measures of both long-term historical volatility (such as historical residual standard deviation) and near-term volatility (such as high-low price ratio, daily standard deviation, and cumulative range over the last 12 months). Other proxies for volatility (log of stock price), corrections for thin trading (serial dependence), and changes in volatility (volume beta) are also included in this descriptor.

2. Momentum

This risk index captures common variation in returns related to recent stock price behavior. Stocks that had positive excess returns in the recent past are grouped separately from those that displayed negative excess returns.

3. Size

This risk index captures differences in stock returns due to differences in the market capitalization of companies.

4. Size Nonlinearity

This risk index captures deviations from linearity in the relationship between returns and log of market capitalization.

5. Trading Activity

This risk index measures the amount of relative trading in each stock. Stocks that are highly traded are likely to be those with greater institutional interest. Such stocks may display different returns behavior compared with those that are not widely held by institutions.

6. Growth

This risk index uses historical growth and profitability measures to predict future earnings growth.

7. Earnings Yield

This risk index combines current and historical earnings-to-price ratios with a measure of analyst-predicted earnings-to-price. Stocks with similar values of earnings yield behave in a similar fashion with respect to their returns.

8. Value

This risk index distinguishes between value stocks and growth stocks using the ratio of book value of equity to market capitalization.

9. Earnings Variability

This risk index measures the variability in earnings and cash flows using both historical measures and analyst predictions.

10. Leverage

This risk index measures the financial leverage of a company.

11. Currency Sensitivity

This risk index measures the sensitivity of a company's stock return to the return on a basket of foreign currencies.

12. Dividend Yield

This risk index computes a measure of predicted dividend yield using the past history of dividends and the market price behavior of the stock.

13. Non-Estimation Universe Indicator

This risk index flags companies outside the estimation universe. It allows the linear factor model to be extended to stocks outside the US-E3 estimation universe.

Descriptor definitions

See Appendix A for a full set of descriptor definitions.

9. US-E3 Industries

Overview

The industry component of the US-E3 model consists of two parts:

- the industry classification scheme (all new for US-E3)
- the industry weights for assets participating in multiple industries (essentially the same as US-E2)

There are 52 industries in the US-E3 classification scheme. The set of industries used is the result of client input and extensive BARRA research. The industry classification scheme used by US-E2, although adequate in many respects, does not capture many relevant features of the current U.S. market. Notable examples are the absence of the software and cellular industries. US-E3 has been constructed to allow for an organic birth/death of industries as the market changes over time.

As with US-E2, intra-asset industry weights are constructed based on valuation-type models using sales, assets, and operating income within industry segments. The Compustat segment data tapes provide the breakdown of these balance sheet items by SIC code.

Industry classification scheme

Mini-industries and industries

After review of tear sheets for all assets in the US-E2 HICAP universe, other industry classification schemes, and in-house micro-industry analysis, we selected 82 mini-industries to begin the asset-by-asset reclassification. Classification of each asset at the mini-industry level has the following benefits:

1. It provides historical information necessary to test and confirm that any mini-industry which could have been placed into one of two or more industries has been correctly positioned within the industry BARRA has selected.

2. It is easy to count the growth of companies within mini-industries. Growth to a "critical mass" can serve as the trigger to consider upgrading a mini-industry to industry status. This vastly aids implementation of an organic classification of industries.

3. Related to 1. and 2. above, we have the relevant information to decide if, given a critical mass of assets within a mini-industry, it is *necessary* to create a new industry. An example of this is Drugs and Biotechnology. Tests of risk forecasts indicate it is not necessary to separate Biotechnology from Drugs, so we have chosen not to do so for US-E3.

Maintaining an ongoing mini-industry classification makes it a relatively simple task to revisit any of the above decisions, should it ever appear necessary. All new assets that appear in the future will be classified at the mini-industry level, and BARRA will be able to add any new mini-industries necessary, should we start seeing companies that do not seem to be appropriately described by our current set of 82.

Sectors

After the US-E3 industries were determined, we created a mapping of industries to 13 sectors. These mappings were based on intuition rather than a data-driven algorithm. For the most part, they should generate little controversy. We have performed tests to confirm that the underlying intuition is borne out by the data. The methodology used was to examine the principal components associated with the industries within each sector. The goal was to determine if the intra-sector industry factor returns "behaved differently" and therefore should not be regarded as being instances of the same sector. This analysis confirmed the current industry-to-sector groupings to be appropriate.

It will be possible to do performance attribution based on sectors with future releases of our Performance products. The time invariance of sectors will make it possible for historical performance across US-E2 and US-E3 historical portfolios. It will also make future introduction of new industries or extinction of current industries simple from a performance point of view. The only requirement is that the sectors remain constant across time. The sectors are sufficiently broad in scope to permit this time invariance.

Industry weights

The methodology necessary to determine industry exposures for companies with more than one business activity is relatively unchanged from US-E2. Briefly, the process is as follows. First, Compustat segment data are used to build what can be regarded as three separate valuation models. Second, the results of each valuation model determine a set of weights, based on fundamental information. Third, the final industry weights are a weighted average of the three weighting schemes. Here is the process in more detail:

1. The valuation models

$$capt_n = \sum_{i=1}^{I} fundamental_{k,n,i} \beta_{k,i}$$

where $fundamental_{k,n,i}$ refers to sales, assets, and operating income of asset n within each industry and $\beta_{k,i}$ is the sensitivity of the nth asset's capitalization to these attributes.

Each of the above three valuation models produces a set of betas that explains the contribution of "$1.00 worth of sales/assets/operating income" to the capitalization of asset n. The estimation of the betas for each model is subject to various controls to prevent unreasonable values.

2. Asset industry weights for each model

Use sales-based valuation as an example:

$$w_{n,sales,i} = \frac{sales_{n,i} \beta_{sales,i}}{\sum_{i=1}^{I} sales_{n,i} \beta_{sales,i}}$$

for the sales-based weight of asset n in industry i.

3. Final weights

We weight each measure into our final industry weights based on reasonable estimates from past model behavior.

Historical assignment of industry and weights

For the purpose of historical estimation, industry assignments from January 1973 to January 1998 were performed on an annual basis. We ran the industry weight program every April during this period (since most U.S. companies end their fiscal year December 31, the beginning of April is when the largest burst of new data would be available). This may result in a loss of timely change of industries for companies that had mergers/splits during the intra-year period. Client feedback should serve to highlight any such misclassifications.

Ongoing assignment of industry and weights

There are three points of note:

1. The weighting program will be run *quarterly* instead of annually. Between quarterly running of the valuation model, the most recent sales, assets, and operating income betas will be used with the most recent available Compustat segment data.

2. We will review the tear sheets of one-twelfth of the assets in the estimation universe each month. The mini-industry associated with each Compustat segment data will be confirmed or corrected. This will ensure that no misclassifications of assets are perpetuated for more than one year. It will also help determine when to remove/add industry exposures resulting from mergers/sell-offs.

3. A well-defined industry override policy is in place. There are two types of overrides:

   - Industry assignments

     Clients are likely to be the source of these. It is quite likely that they know more about what certain companies do than what we can glean from tear sheets and SIC codes. We will assess the client information and make changes to the mini-industry/industry assignments as necessary.

- Weights of assets with multiple industry exposure

  This often arises when clients disagree with the primary industry of a company. In addition to the weighting process described above—call this our *fundamental-based weights*—we will run a style-analysis-based weight program for all companies with multiple-industry exposure. If questions arise about the industry weights, the style-based weights can be compared with the fundamental weights. Without strong support from the style weights, BARRA will stick with the fundamental weights. Clients can be told of our confirming evidence.

Industry evolution

Over time it is reasonable to expect some current industries to disappear and new ones to arise within the U.S. market. This certainly happened over the life cycle of US-E2. Problems associated with the inflexibility of US-E2's industry classification were one of the primary motivating forces for US-E3. The 82 mini-industry classification allows for easy promotion of a current mini-industry to industry status (as well as the less likely recombination of mini-industries into a whole or partially new industry). As long as we have assets identified by their mini-industry, it is easy to keep count of how many of them there are, so once we achieve a critical mass of companies we can consider estimating their factor returns. This also gives us the ability to test to see whether mini-industries behave differently from other mini-industries within the same industry. New mini-industries can be added at any time.

Examples of US-E3 industries that have "been born" since 1975 are:

Industry	Starting Date
Medical Providers & Services	October 1975
Securities & Asset Management	April 1976
Entertainment	November 1980
Computer Software	August 1982
Wireless Telecommunications	June 1989

The starting date for each of the above industries was triggered by an increase in the number of companies whose primary industry was one of the above. The target goal was to have 20 or more assets within each industry.

There are no US-E3 examples of industries that have died out during the January 1975 to current estimation period.

Sector mapping of US-E3 industries

Table 9-1 is an overview of the default sector assignments for US-E3. We have chosen 13 well-defined sectors in order to facilitate stable performance attribution and other model operations as the industry matrix evolves over time. For a more detailed listing of industries, mini-industries, and example companies, *see* Appendix B.

Sector	US-E3 Industry
Basic Materials	Mining & Metals Gold Forest Products & Paper Chemicals
Energy	Energy Reserves & Production Oil Refining Oil Services
Consumer Noncyclicals	Food & Beverages Alcohol Tobacco Home Products Grocery Stores
Consumer Cyclicals	Consumer Durables Motor Vehicles & Parts Apparel & Textiles Clothing Stores Specialty Retail Department Stores Construction & Real Property
Consumer Services	Publishing Media Hotels Restaurants Entertainment Leisure
Industrials	Environmental Services Heavy Electrical Equipment Heavy Machinery Industrial Parts
Utility	Electrical Utilities Gas Utilities
Transport	Railroads Airlines Trucking, Shipping, Air Freight
Health Care	Medical Providers & Services Medical Products Drugs
Technology	Electronic Equipment Semiconductors Computer Hardware & Office Equipment Computer Software Defense & Aerospace
Telecommunications	Telephones Wireless Telecommunications

Table 9-1

Sector Mapping of US-E3 Industries

Sector	US-E3 Industry
Commercial Services	Information Services Industrial Services
Financial	Life & Health Insurance Property & Casualty Insurance Banks Thrifts Securities & Asset Management Financial Services

Table 9-1

Sector Mapping of US-E3 Industries

10. Factor Return Estimation

Overview

US-E3 factor returns for industries and all risk indices *except* Non-Estimation Universe Indicator (NONESTU, formerly LOCAP, which fits the model to smaller assets) are computed using GLS regression over the US-E3 estimation universe. The NONESTU factor return is calculated separately, using the excess return residual to non-NONESTU common factor returns of U.S. assets outside of the estimation universe. The asset weights used for the GLS are ***hsigma***$^{-2}$, where ***hsigma*** is the residual volatility resulting from a simple CAPM regression over the last 60 months of asset returns.

Estimation details

The only screening of data used for computation of the factor returns is the following:

- Assets must have an industry assignment. This removes mutual funds, which had been put in the US-E2 Miscellaneous industry, and any assets which may not have been classified.

- Assets must have a non-missing capitalization. This is done to screen out "dead" assets that have slipped into the historical data.

- Assets must have non-missing hsigma. This is necessary since the asset regression weight is based on this.

- Monthly asset returns must fall within [S&P 500 return - 50%, S&P 500 return + 150%]. This provides protection against possible data errors that have slipped through as well as legitimate large returns that are best prevented from contributing to the factor return estimates. This extreme return screening applies to both the "regular" factor returns and the NONESTU factor.

GLS weights

The US-E3 estimation universe is used for the estimation of industry factor returns and all risk index factor returns *except* the NONESTU factor return.

The regression weight for each asset is $hsigma^{-2}$, truncated to the 99th percentile of all assets in the estimation universe. This is done to prevent an asset with an extraordinarily small hsigma from dominating the entire estimation universe.

Factor returns for historic "newborn" industries

Industries that organically were "born" during the January 1975 to current estimation period have no GLS computed factor returns since there were no assets exposed to them at the time of estimation. A complete time series of factor returns is required for covariance matrix calculation. In these cases we use the time series of GLS factor returns for the parent industry to fill in the January 1975 to starting date factor returns. For example, prior to June 1989 the factor return used in covariance calculation for Wireless Telecommunications will be the same as that which was calculated for Telephones.

The NONESTU factor return

The NONESTU factor return is calculated in the same fashion as the LOCAP factor return in US-E2.

1. Start with all assets *not* in the US-E3 estimation universe.

2. Remove all non-U.S. assets.

3. Exclude all assets associated with extreme returns (see above for definition of "extreme").

4. Compute the specific return of the remaining assets, based on factor returns.

The NONESTU factor return is the capitalization-weighted average of these specific returns.

Testing

Each month we collect performance statistics to allow ongoing monitoring of model performance. These statistics include model R-squared, *t*-statistics, standard errors associated with factor returns, and multicollinearity diagnostics. We test for differences of mini-industries within industries and perform an alternative "thin-industry-adjusted" regression, saving the same model statistics as the ordinary GLS regression.

11. Estimating the Factor Covariance Matrix in US-E3

Overview

We construct the US-E3 common factor covariance matrix using:

- exponential weighting of monthly factor returns with a 90-month half-life

- an extended GARCH (1,1) market volatility forecast to scale the systematic risk portion of the matrix.

US-E2 follows the same methodology, but it uses a plain GARCH(1,1) forecast. *See* Chapter 5. BARRA Multiple-Factor Modeling for details.

12. US-E3 Specific Risk Modeling

Overview

Expect more fluctuating month-to-month specific risk forecasts in US-E3 than in US-E2. This effect will be more pronounced at active risk perspective.

The specific risk model in US-E3 has two important differences from that in US-E2. First, US-E3 has a predictive model for the average level of specific risk. Our research shows that the average level of specific risk displays predictable variation from month to month. As US-E3 incorporates this predictability into its specific risk forecasts, it will be able to predict the average level of specific risk more accurately than US-E2.

The second difference from US-E2 is that US-E3 scales specific risk numbers differently across different size deciles to ensure that forecasts are unbiased within each decile. In contrast, US-E2 uses a single scaling function for all deciles. The finer scaling function used by US-E3 is an improvement over US-E2 and will produce more accurate specific risk forecasts.

These improvements bring our U.S. Equity Model's specific risk forecasting method to the same level as in our other single country equity models. Details are contained in Chapter 5. BARRA Multiple-Factor Modeling.

Appendix A: US-E3 Descriptor Definitions

This Appendix gives the detailed definitions of the descriptors which underlie the risk indices in US-E3. The method of combining these descriptors into risk indices is proprietary to BARRA.

1. Volatility

i) BTSG: Beta times sigma

This is computed as $\sqrt{\beta\sigma_{\varepsilon}}$, where β is the historical beta and σ_{ε} is the historical residual standard deviation. If β is negative, then the descriptor is set equal to zero.

ii) DASTD: Daily standard deviation

This is computed as:

$$\sqrt{N_{days}\left[\sum_{t=1}^{T} w_t r_t^2\right]}$$

where r_t is the return over day t, w_t is the weight for day t, T is the number of days of historical returns data used to compute this descriptor (we set this to 65 days), and N_{days} is the number of trading days in a month (we set this to 23).

iii) HILO: Ratio of high price to low price over the last month

This is calculated as:

$$\log\left(\frac{P_H}{P_L}\right)$$

where P_H and P_L are the maximum price and minimum price attained over the last one month.

iv) LPRI: Log of stock price

This is the log of the stock price at the end of last month.

v) CMRA: Cumulative range

Let Z_t be defined as follows:

$$Z_t = \sum_{s=1}^{t} \log(1 + r_{i,s}) - \sum_{s=1}^{t} \log(1 + r_{f,s})$$

where $r_{i,s}$ is the return on stock i in month s, and $r_{f,s}$ is the risk-free rate for month s. In other words, Z_t is the cumulative return of the stock over the risk-free rate at the end of month t. Define $Z_{\max}$ and $Z_{\min}$ as the maximum and minimum values of Z_t over the last 12 months. *CMRA* is computed as:

$$\log\left(\frac{1 + Z_{\max}}{1 + Z_{\min}}\right)$$

vi) VOLBT: Sensitivity of changes in trading volume to changes in aggregate trading volume

This may be estimated by the following regression:

$$\frac{\Delta V_{i,t}}{N_{i,t}} = a + b\frac{\Delta V_{M,t}}{N_{M,t}} + \xi_{i,t}$$

where $\Delta V_{i,t}$ is the change in share volume of stock i from week t-1 to week t, $N_{i,t}$ is the *average* number of shares outstanding for stock i at the beginning of week t-1 and week t, $\Delta V_{M,t}$ is the change in volume on the aggregate market from week t-1 to week t, and $N_{M,t}$ is the average number of shares outstanding for the aggregate market at the beginning of week t-1 and week t.

vii) SERDP: Serial dependence

This measure is designed to capture serial dependence in residuals from the market model regressions. It is computed as follows:

$$SERDP = \frac{\frac{1}{T-2}\sum_{t=3}^{T}(e_t + e_{t-1} + e_{t-2})^2}{\frac{1}{T-2}\sum_{t=3}^{T}(e_t^2 + e_{t-1}^2 + e_{t-2}^2)}$$

where e_t is the residual from the market model regression in month t, and T is the number of months over which this regression is run (typically, T = 60 months).

viii) OPSTD: Option-implied standard deviation

This descriptor is computed as the implied standard deviation from the Black-Scholes option pricing formula using the price on the closest to at-the-money call option that trades on the underlying stock.

2. Momentum

i) RSTR: Relative strength

This is computed as the cumulative excess return (using continuously compounded monthly returns) over the last 12 months—i.e.,

$$RSTR = \sum_{t=1}^{T}\log(1 + r_{i,t}) - \sum_{t=1}^{T}\log(1 + r_{f,t})$$

where $r_{i,t}$ is the arithmetic return of the stock in month i, and $r_{f,t}$ is the arithmetic risk-free rate for month i. This measure is usually computed over the last one year—i.e., T is set equal to 12 months.

ii) HALPHA: Historical alpha

This descriptor is equal to the alpha term (i.e., the intercept term) from a 60-month regression of the stock's excess returns on the S&P 500 excess returns.

3. Size

i) LNCAP: Log of market capitalization

This descriptor is computed as the log of the market capitalization of equity (price times number of shares outstanding) for the company.

4. Size Nonlinearity

i) LCAPCB: Cube of the log of market capitalization

This risk index is computed as the cube of the normalized log of market capitalization.

5. Trading Activity

i) STOA: Share turnover over the last year

STOA is the annualized share turnover rate using data from the last 12 months—i.e., it is equal to $V_{ann}/\bar{N}_{out}$, where V_{ann} is the total trading volume (in number of shares) over the last 12 months and $\bar{N}_{out}$ is the average number of shares outstanding over the previous 12 months (i.e., it is equal to the average value of the number of shares outstanding at the beginning of each month over the previous 12 months).

ii) STOQ: Share turnover over the last quarter

This is computed as the annualized share turnover rate using data from the most recent quarter. Let V_q be the total trading volume (in number of shares) over the most recent quarter and let $\bar{N}_{out}$ be the average number of shares outstanding over the period (i.e., $\bar{N}_{out}$ is equal to the average value of the number of shares outstanding at the beginning of each month over the previous three months). Then, *STOQ* is computed as $4V_q/\bar{N}_{out}$.

iii) STOM: Share turnover over the last month

This is computed as the share turnover rate using data from the most recent month—i.e., it is equal to the number of shares traded last month divided by the number of shares outstanding at the beginning of the month.

iv) STO5: Share turnover over the last five years

This is equal to the annualized share turnover rate using data from the last 60 months. In symbols, *STO5* is given by:

$$STO5 = \frac{12\left[\frac{1}{T}\sum_{s=1}^{T} V_s\right]}{\overline{N}_{out}}$$

where V_s is equal to the total trading volume in month s and $\overline{N}_{out}$ is the average number of shares outstanding over the last 60 months.

v) FSPLIT: Indicator for forward split

This descriptor is a 0-1 indicator variable to capture the occurrence of forward splits in the company's stock over the last two years.

vi) VLVR: Volume to variance

This measure is calculated as follows:

$$VLVR = \log\left(\frac{\frac{12}{T}\sum_{s=1}^{T} V_s P_s}{\sigma_\varepsilon}\right)$$

where V_s equals the number of shares traded in month s, P_s is the closing price of the stock at the end of month s, and σ_ε is the estimated residual standard deviation. The sum in the numerator is computed over the last 12 months.

i) PAYO: Payout ratio over five years

This measure is computed as follows:

$$PAYO = \frac{\frac{1}{T}\sum_{t=1}^{T} D_t}{\frac{1}{T}\sum_{t=1}^{T} E_t}$$

where D_t is the aggregate dividend paid out in year t and E_t is the total earnings available for common shareholders in year t. This descriptor is computed using the last five years of data on dividends and earnings.

ii) VCAP: Variability in capital structure

This descriptor is measured as follows:

$$VCAP = \frac{\frac{1}{T-1}\sum_{t=2}^{T}\left(\left|N_{t-1} - N_t\right|P_{t-1} + \left|LD_{t-1} - LD_t\right| + \left|PE_t - PE_{t-1}\right|\right)}{CE_T + LD_T + PE_T}$$

where N_{t-1} is the number of shares outstanding at the end of time t-1; P_{t-1} is the price per share at the end of time t-1; LD_{t-1} is the book value of long-term debt at the end of time period t-1; PE_{t-1} is the book value of preferred equity at the end of time period t-1; and $CE_T + LD_T + PE_T$ are the book values of common equity, long-term debt, and preferred equity as of the most recent fiscal year.

iii) AGRO: Growth rate in total assets

To compute this descriptor, the following regression is run:

$$TA_{it} = a + bt + \zeta_{it}$$

where TA_{it} is the total assets of the company as of the end of year t, and the regression is run for the period =1,...,5. *AGRO* is computed as follows:

$$AGRO = \frac{b}{\frac{1}{T}\sum_{t=1}^{T} TA_{it}}$$

where the denominator average is computed over all the data used in the regression.

iv) EGRO: Earnings growth rate over last five years

First, the following regression is run:

$$EPS_t = a + bt + \zeta_t$$

where EPS_t is the earnings per share for year t. This regression is run for the period t=1,...,5. *EGRO* is computed as follows:

$$EGRO = \frac{b}{\frac{1}{T}\sum_{t=1}^{T} EPS_t}$$

v) EGIBS: Analyst-predicted earnings growth

This is computed as follows:

$$EGIBS = \frac{(EARN - EPS)}{(EARN + EPS)/2}$$

where *EARN* is a weighted average of the median earnings predictions by analysts for the current year and next year, and *EPS* is the sum of the four most recent quarterly earnings per share.

vi) DELE: Recent earnings change

This is a measure of recent earnings growth and is measured as follows:

$$DELE = \frac{(EPS_t - EPS_{t-1})}{(EPS_t + EPS_{t-1})/2}$$

where EPS_t is the earnings per share for the most recent year, and EPS_{t-1} is the earnings per share for the previous year. We set this to missing if the denominator is non-positive.

7. Earnings Yield

i) EPIBS: Analyst-predicted earnings-to-price

This is computed as the weighted average of analysts' median predicted earnings for the current fiscal year and next fiscal year divided by the most recent price.

ii) ETOP: Trailing annual earnings-to-price

This is computed as the sum of the four most recent quarterly earnings per share divided by the most recent price.

iii) ETP5: Historical earnings-to-price

This is computed as follows:

$$ETP5 = \frac{\frac{1}{T}\sum_{t=1}^{T} EPS_t}{\frac{1}{T}\sum_{t=1}^{T} P_t}$$

where EPS_t is equal to the earnings per share over year t, and P_t is equal to the closing price per share at the end of year t.

8. Value

i) BTOP: Book-to-price ratio

This is the book value of common equity as of the most recent fiscal year end divided by the most recent value of the market capitalization of the equity.

9. Earnings Variability

i) VERN: Variability in earnings

This measure is computed as follows:

$$VERN = \frac{\left(\frac{1}{T-1}\sum_{t=1}^{T}\left(E_t - \overline{E}\right)^2\right)^{1/2}}{\frac{1}{T}\sum_{t=1}^{T}E_t}$$

where E_t is the earnings at time t (t=1,...,5) and $\overline{E}$ is the average earnings over the last five years. *VERN* is the coefficient of variation of earnings.

ii) VFLO: Variability in cash flows

This measure is computed as the coefficient of variation of cash flow using data over the last five years—i.e., it is computed in an identical manner to *VERN*, with cash flow being used in place of earnings. Cash flow is computed as earnings plus depreciation plus deferred taxes.

iii) EXTE: Extraordinary items in earnings

This is computed as follows:

$$EXTE = \frac{\frac{1}{T}\sum_{t=1}^{T}\left|EX_t + NRI_t\right|}{\frac{1}{T}\sum_{t=1}^{T}E_t}$$

where EX_t is the value of extraordinary items and discontinued operations, NRI_t is the value of non-operating income, and E_t is the earnings available to common before extraordinary items. The descriptor uses data over the last five years.

iv) SPIBS: Standard deviation of analysts' prediction to price

This is computed as the weighted average of the standard deviation of IBES analysts' forecasts of the firm's earnings per share for the current fiscal year and next fiscal year divided by the most recent price.

10. Leverage

i) MLEV: Market leverage

This measure is computed as follows:

$$MLEV = \frac{ME_t + PE_t + LD_t}{ME_t}$$

where ME_t is the market value of common equity, PE_t is the book value of preferred equity, and LD_t is the book value of long-term debt. The value of preferred equity and long-term debt are as of the end of the most recent fiscal year. The market value of equity is computed using the most recent month's closing price of the stock.

ii) BLEV: Book leverage

This measure is computed as follows:

$$BLEV = \frac{CEQ_t + PE_t + LD_t}{CEQ_t}$$

where CEQ_t is the book value of common equity, PE_t is the book value of preferred equity, and LD_t is the book value of long-term debt. All values are as of the end of the most recent fiscal year.

iii) DTOA: Debt-to-assets ratio

This ratio is computed as follows:

$$DTOA = \frac{LD_t + DCL_t}{TA_t}$$

where LD_t is the book value of long-term debt, DCL_t is the value of debt in current liabilities, and TA_t is the book value of total assets. All values are as of the end of the most recent fiscal year.

iv) SNRRT: Senior debt rating

This descriptor is constructed as a multi-level indicator variable of the debt rating of a company.

11. Currency Sensitivity

i) CURSENS: Exposure to foreign currencies

To construct this descriptor, the following regression is run:

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}$$

where r_{it} is the excess return on the stock and r_{mt} is the excess return on the S&P 500 Index. Let ε_{it} denote the residual returns from this regression. These residual returns are in turn regressed against the contemporaneous and lagged returns on a basket of foreign currencies, as follows:

$$\varepsilon_{it} = c_i + \gamma_{i1}(FX)_t + \gamma_{i2}(FX)_{t-1} + \gamma_{i3}(FX)_{t-2} + u_{it}$$

where ε_{it} is the residual return on stock i, $(FX)_t$ is the return on an index of foreign currencies over month t, $(FX)_{t-1}$ is the return on the same index of foreign currencies over month t-1, and $(FX)_{t-2}$ is the return on the same index over month t-2. The risk index is computed as the sum of the slope coefficients γ_{i1}, γ_{i2}, and γ_{i3}—i.e., $CURSENS = \gamma_{i1} + \gamma_{i2} + \gamma_{i3}$.

12. Dividend Yield

i) P_DYLD: Predicted dividend yield

This descriptor uses the last four quarterly dividends paid out by the company along with the returns on the company's stock and future dividend announcements made by the company to come up with a BARRA-predicted dividend yield.

13. Non-Estimation Universe Indicator

i) NONESTU: Indicator for firms outside US-E3 estimation universe

This is a 0-1 indicator variable: It is equal to 0 if the company is in the US-E3 estimation universe and equal to 1 if the company is outside the US-E3 estimation universe.

Appendix B: US-E3 Industries, Mini-Industries, Example Companies, and Codes

Below you will find definitions and example companies for the US-E3 industry definitions. US-E3 industry definitions consist of two types: (1) main industries and (2) mini-industries. Main industries and their codes are indicated by **boldface** type in the table on the following pages. Each main US-E3 industry is comprised of one or more mini-industries; if more than one, secondary mini-industries are listed beneath the main industry. For example, the **Forest Products & Paper** main industry consists of three mini-industries: **Forest Products & Paper**; Household Paper; and Office Paper. The mini-industry which dominates the main industry (in this example, **Forest Products & Paper**) will usually determine the name of the main industry.

Mini-industries serve two purposes. Some are being watched to determine whether they will grow into main industries at some future date—e.g., Gaming and Biotechnology. Others have been classified separately to allow for future testing and confirmation that the clusters of mini-industries have been correctly aggregated into main industries.

The industries are aggregated into sectors, which provide a coarser breakdown of business activities. Sectors can be used for display purposes in US-E3 but play no role in estimation.

Sector	US-E3 Industry Definition	Code
Basic Materials	**Mining & Metals** Aluminum, coal, iron and steel, stainless steel, steel castings, copper, beryllium, nickel, titanium, uranium. Metal and glass containers. Metal recycling. Does not include aluminum foil for households. *Examples:* Alcoa, Bethlehem Steel	**MINING**
	Gold Gold and silver. *Example:* Homestake Mining	**GOLD**
	Forest Products & Paper Lumber, wood, paper, newsprint, paper and cardboard containers, paper machine clothing (e.g., conveyor belts). *Example:* Georgia Pacific	**FOREST**
	Household Paper Toilet paper, paper towels, tissues. Household paper wholesaling. *Example:* Kimberly-Clark	HSPAPER
	Office Paper Stationery, office paper supplies. *Example:* Nashua	OFPAPER
	Chemicals Chemicals, paints, plastics, plastic containers, coatings, gases, adhesives, inks, fibers. Does not include household chemicals and plastics. *Example:* Dow Chemical	**CHEM**
	Fertilizers Fertilizers, agricultural chemicals (pesticides). *Example:* Monsanto	FERT
Energy	**Energy Reserves & Production** Oil and gas exploration, reserves, and production. *Example:* Exxon	**RESERVES**
	Oil Refining Oil refining and marketing. *Example:* Quaker State	**OILREF**
	Gas Pipelines Gas pipelines and distribution. *Example:* Williams Companies	GASPIPE

Sector	US-E3 Industry Definition	Code
Energy (continued)	**Oil Services** Oil drilling, oil platform operating, oil platform support services, drill bits, drilling tools. *Examples:* Parker Drilling, Schlumberger	**OILSERV**
Consumer Noncyclicals	**Food & Beverages** Food companies (processed food), raw agriculture products, beverage companies, soda bottling, flower-growing, veterinary services. Does not include fertilizer. *Examples:* Kellogg, Tyson Foods, Coca-Cola	**FOODBEV**
	Food Wholesale Food distribution and wholesaling. *Example:* Fleming Companies	WHOLEFD
	Alcohol Beer, wine, and spirits. Alcohol wholesaling. *Example:* Anheuser-Busch	**ALCOHOL**
	Tobacco Cigars, cigarettes, pipe tobacco, leaf tobacco dealers, chewing tobacco, tobacco distribution. *Example:* Philip Morris	**TOBACCO**
	Home Products Soaps, housewares, cosmetics, personal care, skin care, beauty care, dental care, household chemicals, household plastics. *Example:* Procter & Gamble	**HOMEPROD**
	Grocery Stores Grocery stores. *Example:* Safeway	**RETFOOD**
Consumer Cyclicals	**Consumer Durables** Home furniture, appliances, lawn mowers, snow blowers, televisions, floor coverings, non-textile home furnishings, luggage, cutlery, china. Consumer durable wholesaling. *Examples:* Maytag, Sunbeam, Black & Decker	**DURABLES**
	Office Furniture Office Furniture. *Example:* Hon Industries	OFFURN

Sector	US-E3 Industry Definition	Code
Consumer Cyclicals (continued)	**Motor Vehicles & Parts** Car and auto part manufacturing. Car batteries. Not auto part retailing. *Examples:* Ford, GM	**CARS**
	Trucks Truck manufacturing. Not heavy equipment (e.g., cranes). *Example:* Navistar International Corp.	TRUCKS
	Motor Homes Motor homes and trailers. *Example:* Winnebago	MOTORHM
	Apparel, Textiles Manufacturing of apparel, textiles, shoes, textile home furnishings (towels, sheets, etc.), processing of fibers for textiles, wholesale but not retail sale of apparel. *Examples:* Liz Claiborne, Nike	**APPAREL**
	Clothing Stores Specialty apparel retailers. Does not include fabric stores. *Examples:* Gap, Nordstrom	**RETAPP**
	Specialty Retail Sells one type of item or uses one concept (e.g., items under $10). Not apparel retailers. Includes retail auto supply stores, consumer electronics stores, fabric stores, catalog marketing, telemarketing, auctioneers. *Examples:* Home Depot, Circuit City	**RETSPEC**
	Drug Stores *Example:* Longs Drugs	DRUGSTOR
	Department Stores Retailers who sell widely diverse products, department stores. *Examples:* Dayton Hudson, Wal-Mart	**RETDIV**
	Construction & Real Property Building materials, residential lighting and fixtures, home builders, building managers, equity REITs. Wholesale construction materials. *Example:* Kaufman & Broad	**CONST**
	Concrete Concrete and construction aggregates *Example:* Lone Star Industries, Inc.	CONCRETE
	Manufactured Housing and Mobile Homes *Example:* Skyline Corp.	MANHOUSE

Sector	US-E3 Industry Definition	Code
Consumer Cyclicals (continued)	Engineering Engineering and construction firms, industrial construction, plant construction. *Example:* Stone & Webster	ENGNR
Consumer Services	**Publishing** Newspaper and magazine publishers, book publishers, greeting cards. *Examples:* Dow Jones, New York Times	**PUB**
	Commercial Printing Check and business form printing. *Example:* R.R. Donnelley	PRINTING
	Media Radio and TV stations, cable TV stations. Not TV programming. *Example:* CBS	**MEDIA**
	Satellite Communications *Example:* COMSAT	SATELLTE
	Hotels Hotels and motels. *Example:* Hilton Hotels	**HOTEL**
	Gaming Casinos. *Example:* Caesars World	GAMING
	Restaurants *Example:* McDonald's	**RESTRNT**
	Entertainment Movies, TV programming, theaters, theme parks, cruises. *Example:* Disney	**ENT**
	Leisure Golf clubs, boats, toys, photography, entertainment systems, bicycles, novelties, camps, recreational vehicle parks, jewelry. *Example:* Eastman Kodak	**LEISURE**
	Gaming Equipment *Example:* International Game	GAMEQUIP
	Motorcycles *Example:* Harley-Davidson	MCYCLE

Sector	US-E3 Industry Definition	Code
Industrials	**Environmental Services** Waste management, hazardous materials disposal, cogeneration and independent power. Environmental consulting. *Example:* Browning-Ferris Industries (BFI)	**ENSERV**
	Heavy Electrical Equipment Electrical equipment, electric power generation equipment, cables, wire, insulation, connectors. Not electronics, batteries, elevators. *Examples:* GE, Westinghouse	**ELECTRIC**
	Heavy Machinery Tractors, cranes, fire trucks, sweeper machines. *Example:* Caterpillar	**HMACHINE**
	Industrial Parts Industrial manufacture of bearings, gears, pumps, equipment, batteries. Elevators. *Examples:* Applied Industrial Technologies, Inc., Timken	**INDPARTS**
	Machine Tools Manufacturing of machine tools. Heavy industry machine tools. *Example:* Cincinnati Milacron	MTOOLS
	Hand Tools Manufacturing of hand tools. *Example:* Snap-On Tools	HTOOLS
	Machinery Manufacturing of machinery, engines. *Example:* Briggs & Stratton	MACHINE
Utilities	**Electric Utilities** *Example:* Pacific Gas & Electric	**ELECUTIL**
	Gas Utilities *Example:* Brooklyn Union Gas	**GASWATER**
	Water Utilities *Example:* American Water Works	WATUTIL
Transport	**Railroads** Railroads and rolling stock leasing. *Example:* Burlington Northern	**RAILROAD**
	Airlines Airlines and airport ground services. Not air freight. *Examples:* AMR, Delta	**AIRLINE**

Sector	US-E3 Industry Definition	Code
Transport (continued)	**Trucking, Shipping, Air Freight** Sea, land freight. *Examples:* CNF Transportation, Inc., Alexander & Baldwin	**FREIGHT**
	Air Freight *Example:* American Express	AIRFRGHT
Health Care	**Medical Providers & Services** Hospitals, nursing homes, surgical centers, HMOs, rehabilitation providers, laboratories, hospital pharmacy management, apartments with available nursing, cemetery and funeral homes, medical equipment rental. *Example:* Humana	**MEDPROV**
	Medical Products High-tech and low-tech medical products.	**MEDPROD**
	Medical Technology X-ray machines, ultrasound, CAT scan, angiographic products, implants, pacemakers. *Examples:* Advanced Technology, Boston Scientific	MEDTECH
	Medical Supplies Band-Aids, catheters, needles, knives, blood collection vials. *Example:* Kendall	MEDSUPP
	Drugs Drug production via traditional chemical processes. *Examples:* Merck, Eli Lilly	**DRUGS**
	Biotechnology Drug production via DNA technology, monoclonal antibodies. *Examples:* Genentech, Amgen	BIOTECH

Sector	US-E3 Industry Definition	Code
Technology	**Electronic Equipment** Analytical instruments, instrumentation, test and measurement electronics, high precision motors, electronic power supplies, electronic controls, circuit board manufacturing, antennas, electrical switches, circuit protectors, sensors, thermostats, security systems, membrane separation technology. *Examples:* Tektronix, Varian	**MANHT**
	Communications Equipment Equipment for digital switching, telecommunications, visual telecommunications, video transmission and broadcasting, voice and data networking, voice mail and electronic mail, and telephones; optical fiber manufacturing, paging equipment. *Examples:* Oak Industries, ADC Telecommunications	COMMEQP
	Semiconductors Manufacture of semiconductors, chips, equipment for semiconductor manufacturing. *Example:* Intel	**SEMICOND**
	Computer Hardware & Office Equipment Computers, disk and tape drives, network equipment, copiers, office machines, bar code scanners, credit card verification equipment, automatic toll collectors, and automatic teller machines. *Example:* IBM	**COMPUTER**
	Computer Software *Example:* Microsoft	**SOFTWARE**
	Defense & Aerospace Manufacturing for defense, including aircraft, tanks, submarines, and supplies. Civil aeronautics, space exploration. Aircraft parts. *Examples:* Lockheed, Boeing, General Dynamics	**DEFAERO**

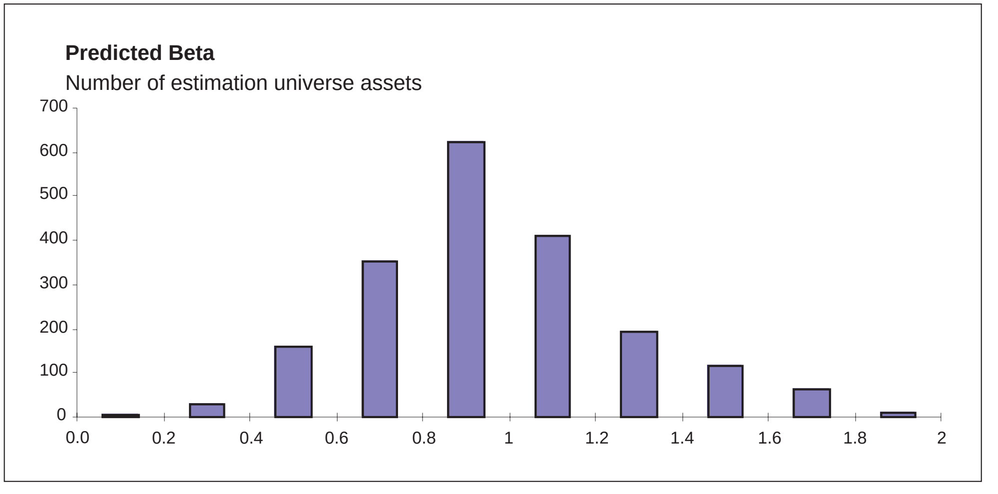
Sector	US-E3 Industry Definition	Code
Telecommunications	**Telephones** Long-distance companies. *Example:* AT&T	**PHONE**
	Local Phone Companies Local phone companies and Baby Bells. *Examples:* NYNEX, Pacific Telesis	LOCPHONE
	Wireless Telecommunications Cellular phones and pagers. *Examples:* GTE Contel, Nextel Communications	**CELLULAR**
Commercial Services	**Information Services** Account keeping, payroll processing, news feeds, user-searchable databases, research-based financial information, translation services, computer centers, market research, management consulting, advertising, electronic publishing. Travel agents. *Examples:* Automatic Data Processing, Dun & Bradstreet	**INFOSERV**
	Industrial Services Janitorial and housekeeping services, plumbing, pest control, employment agencies, office temps, uniform rental, truck and auto fleet management, relocation services, employee and student education, training, and testing; security services, industrial plant management, car and truck leasing. Auto repair. *Example:* Manpower Inc.	**INDSERV**
Financial	**Life and Health Insurance** Includes life insurance, annuities, health insurance, disability insurance. Not HMOs. *Examples:* Equitable, UNUM	**LIFEINS**
	Property and Casualty Insurance Property insurance, casualty insurance, municipal bond insurance, title insurance, liability insurance, workers' compensation insurance. Not HMOs. *Example:* Allstate	**OTHERINS**
	Banks Local and regional banks. *Example:* BayBank	**BANK**
	Money Center Banks *Example:* J.P. Morgan	MONCENTR
	Thrifts *Example:* Dime Savings Bank	**THRIFT**

Sector	US-E3 Industry Definition	Code
Financial (continued)	**Securities & Asset Management** Brokers, mutual fund companies, asset management companies. *Example:* Merrill Lynch	**SECASSET**
	Financial Services Mortgage brokers, consumer loans, car loans, student loans, home loans, credit cards, tax preparation, loan agencies of the U.S. government; also credit card processing, check guarantee, loan guarantee and collection, wire transfer, mortgage REITs. Credit unions, pawnshops, patents. *Example:* FNMA	**FINSERV**
	Insurance Services and Brokers Insurance reporting and consulting, insurance claims adjusting, insurance brokers. *Example:* Marsh & McClennan	INSERVBR

Appendix C: US-E3 Frequency Distributions
for Predicted Beta, Specific Risk, Risk Indices

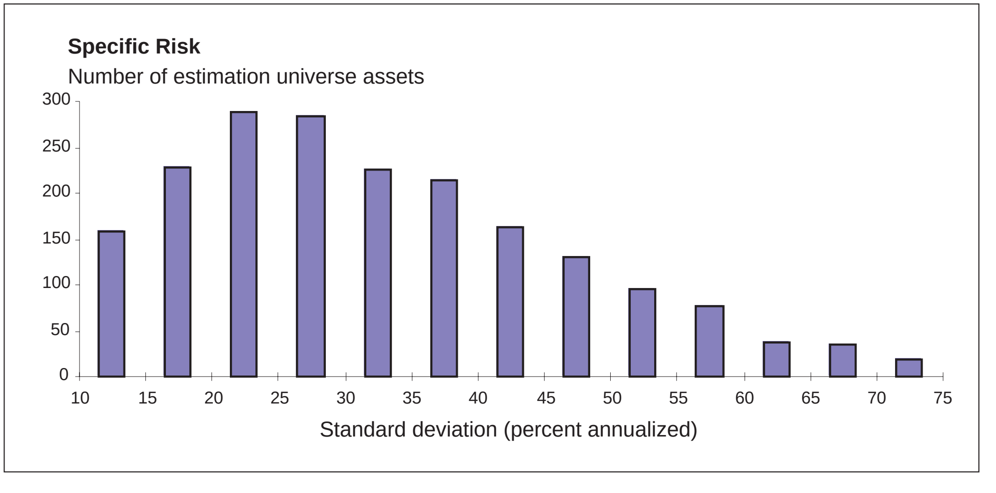
Predicted beta

Predicted US-E3 beta frequency distribution of 1,941 estimation universe assets as of December 1997:



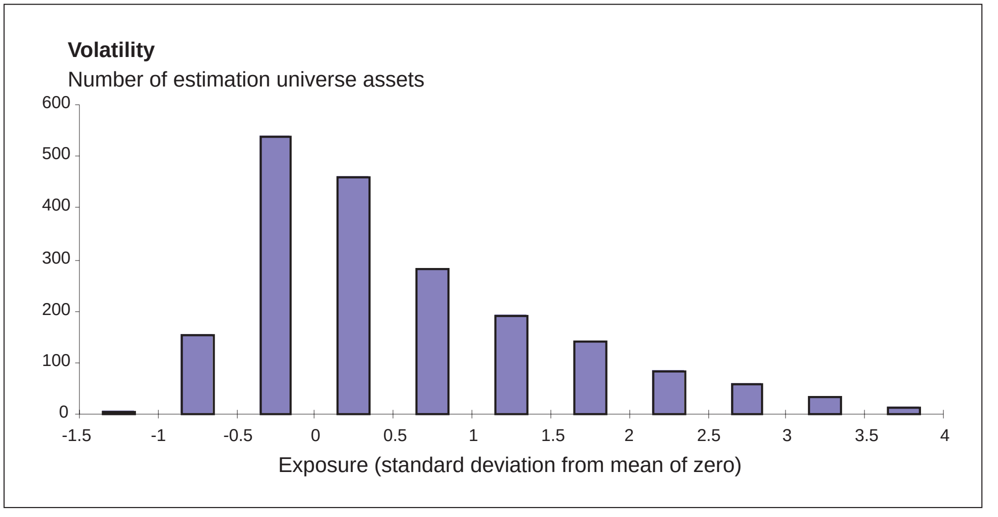
Specific risk

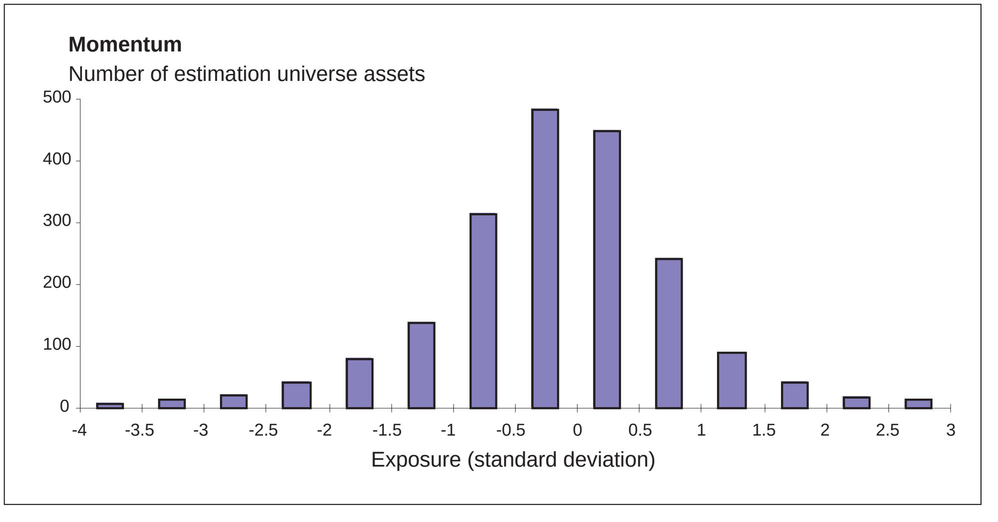
Predicted US-E3 specific risk frequency distribution of 1,941 estimation universe assets as of December 1997:

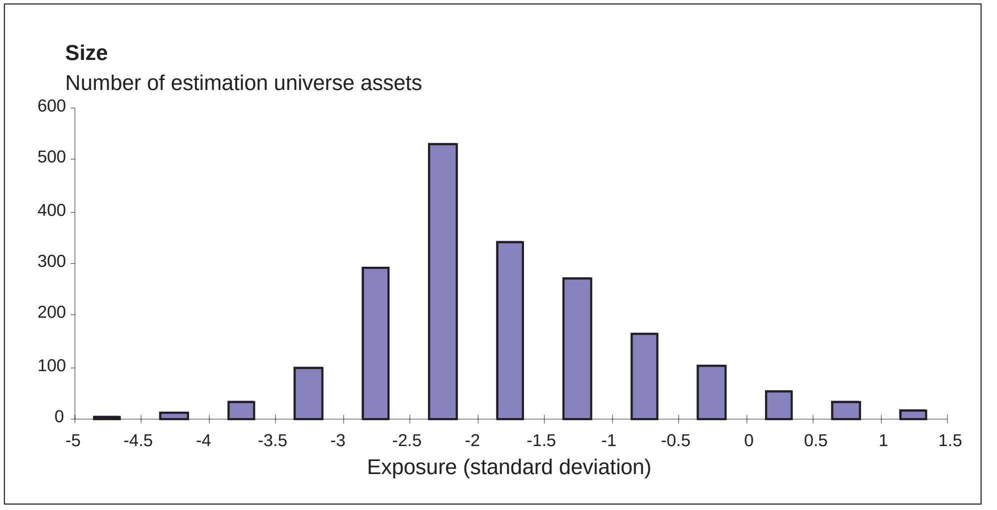


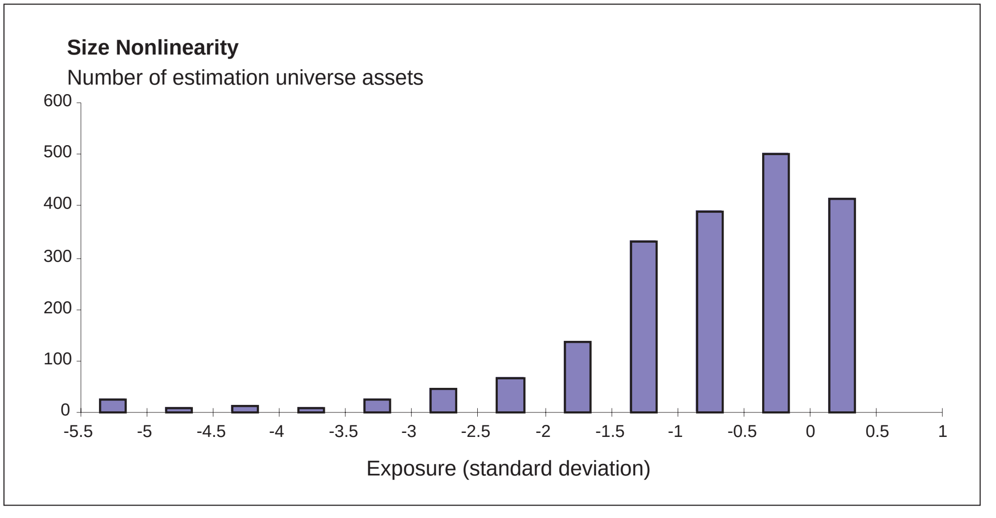
Risk indices

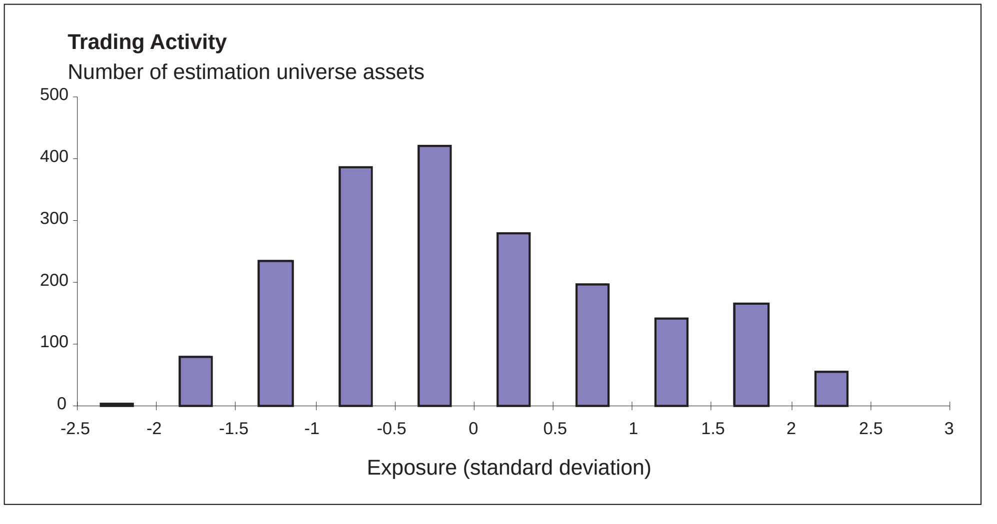
US-E3 risk index frequency distributions of 1,941 estimation universe assets as of December 1997:

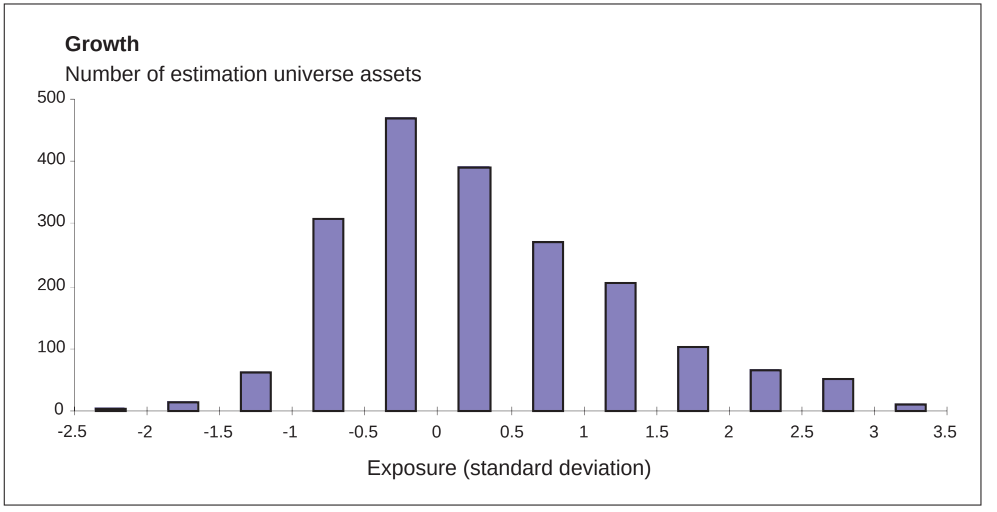


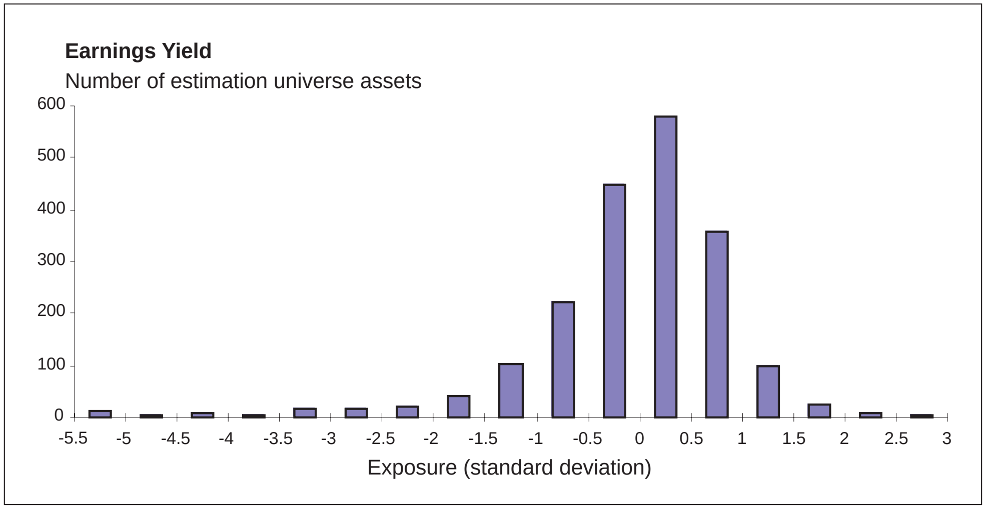


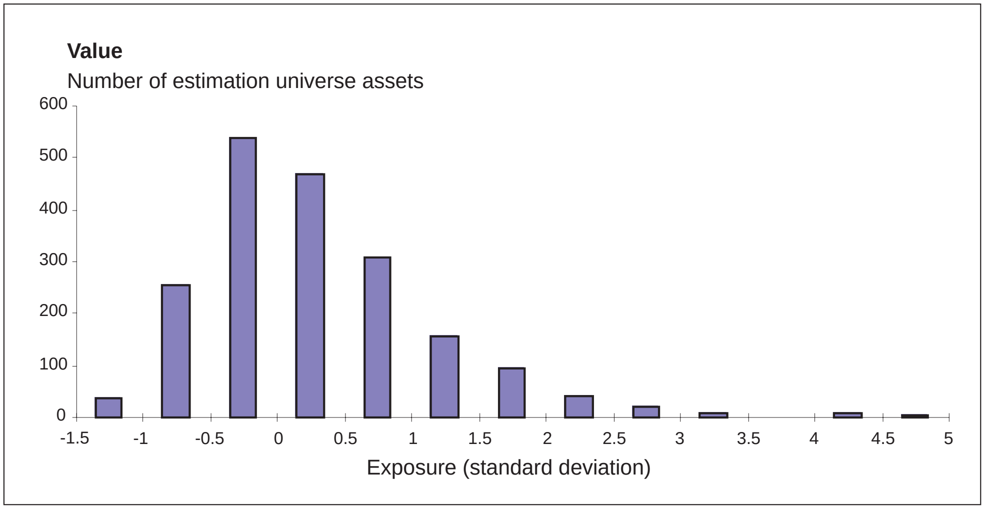

Size
Number of estimation universe assets
600
500
400
300
200
100
0
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
Exposure (standard deviation)

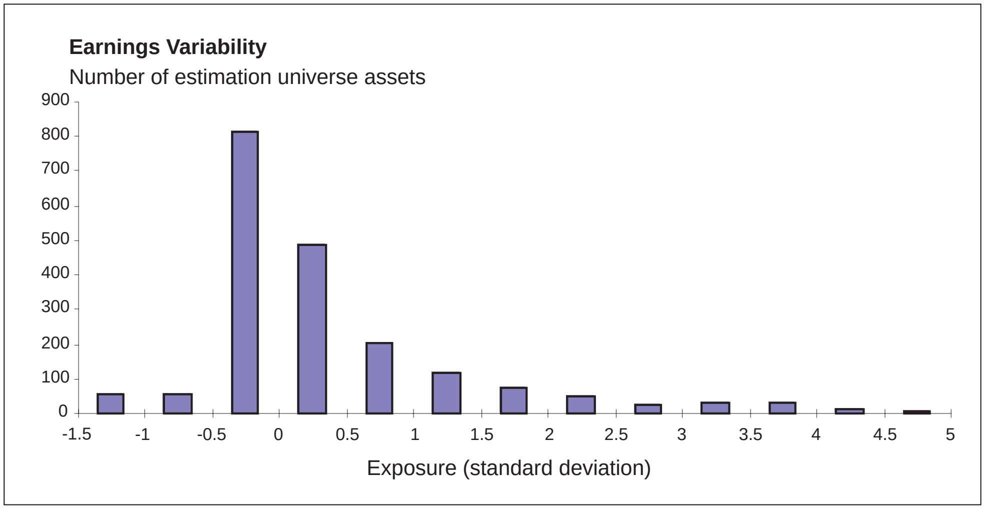

Size Nonlinearity
Number of estimation universe assets
600
500
400
300
200
100
0
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
Exposure (standard deviation)

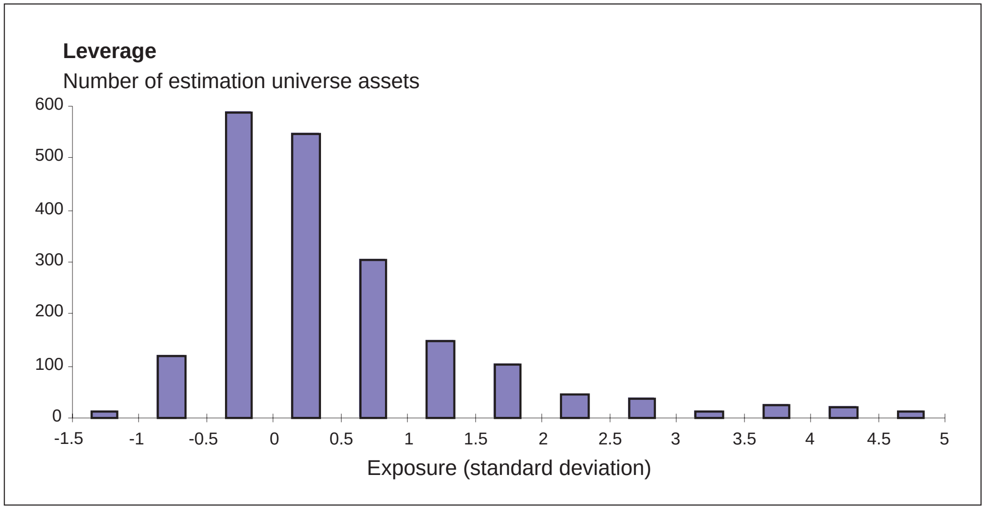

Trading Activity
Number of estimation universe assets
500
400
300
200
100
0
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Exposure (standard deviation)

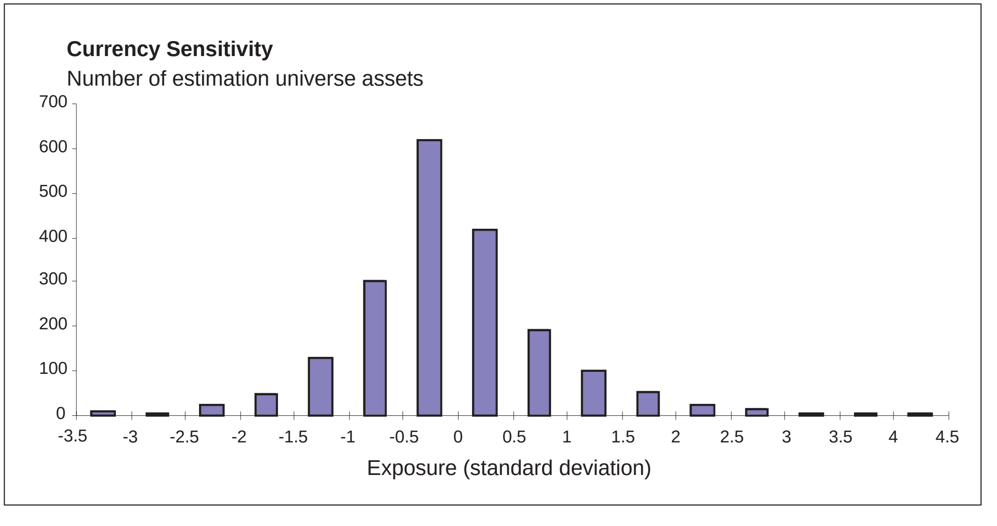

Growth
Number of estimation universe assets
500
400
300
200
100
0
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
Exposure (standard deviation)

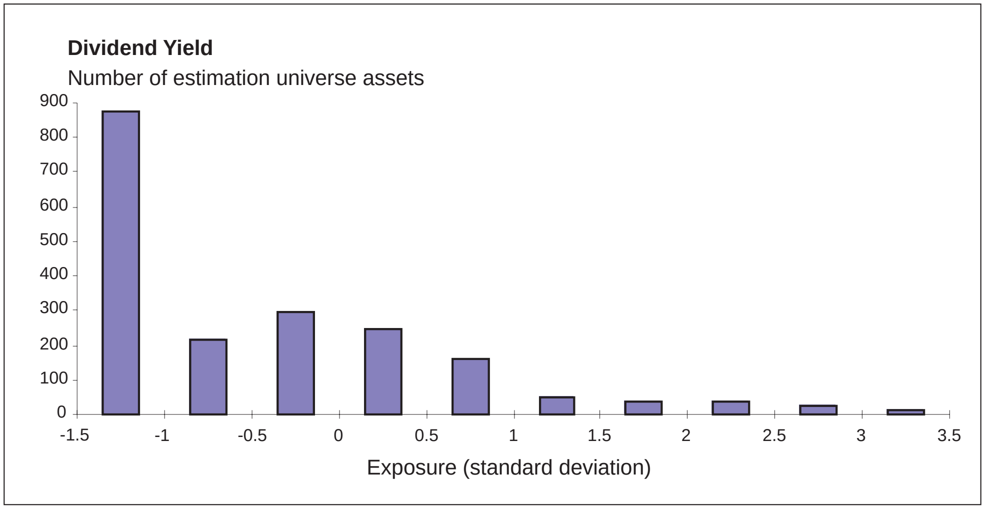

Earnings Yield
Number of estimation universe assets
600
500
400
300
200
100
0
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Exposure (standard deviation)


Value
Number of estimation universe assets
600
500
400
300
200
100
0
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Exposure (standard deviation)


Earnings Variability
Number of estimation universe assets
900
800
700
600
500
400
300
200
100
0
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Exposure (standard deviation)

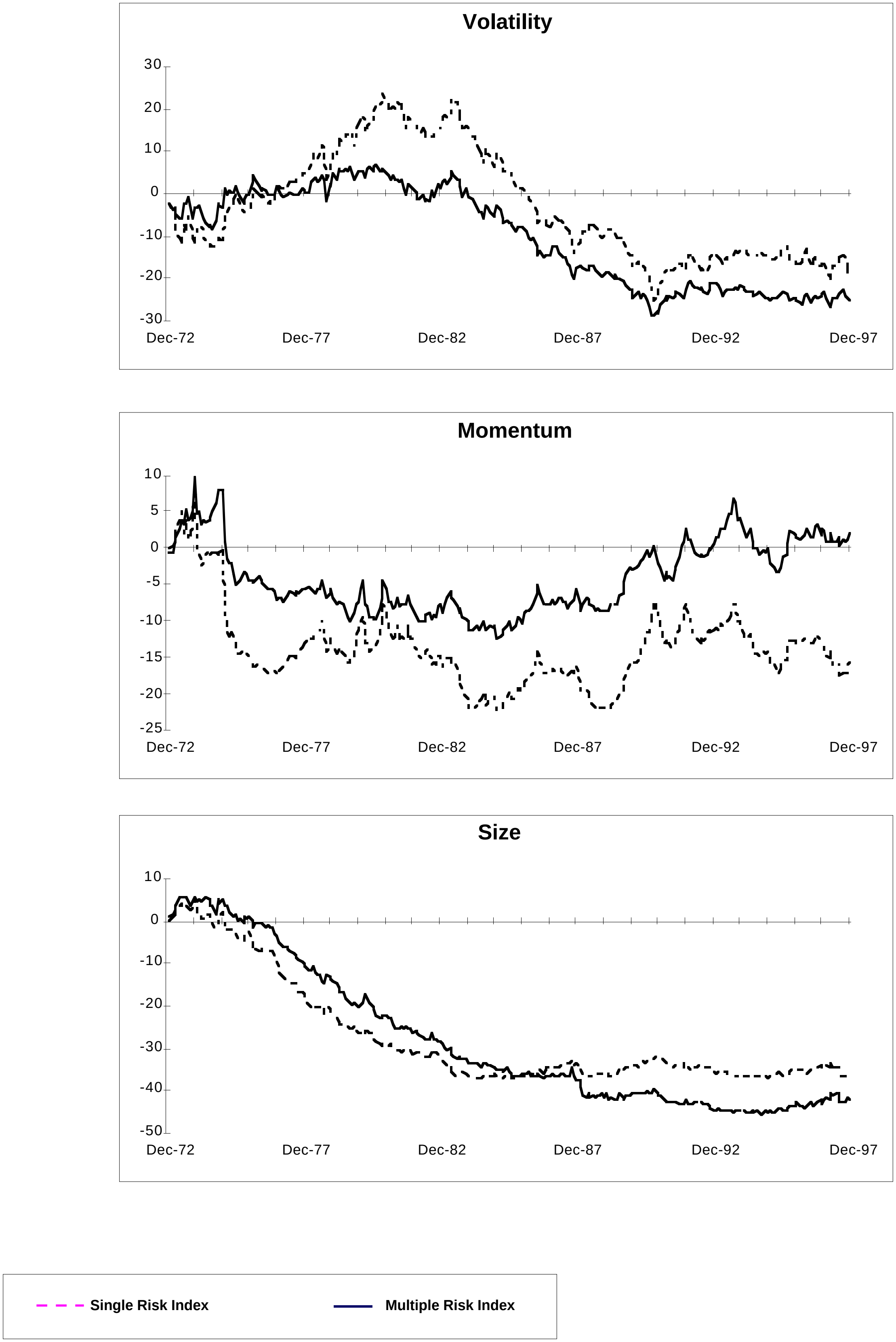

Leverage
Number of estimation universe assets
600
500
400
300
200
100
0
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Exposure (standard deviation)

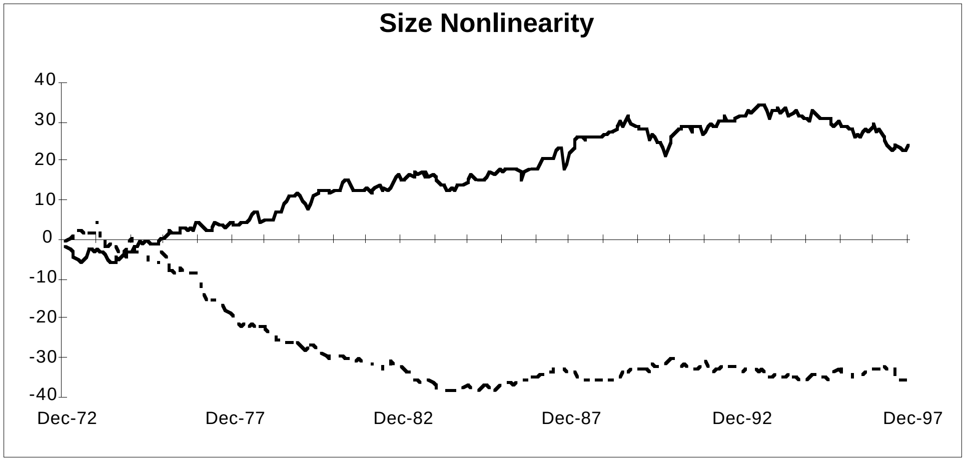

Currency Sensitivity
Number of estimation universe assets
700
600
500
400
300
200
100
0
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Exposure (standard deviation)

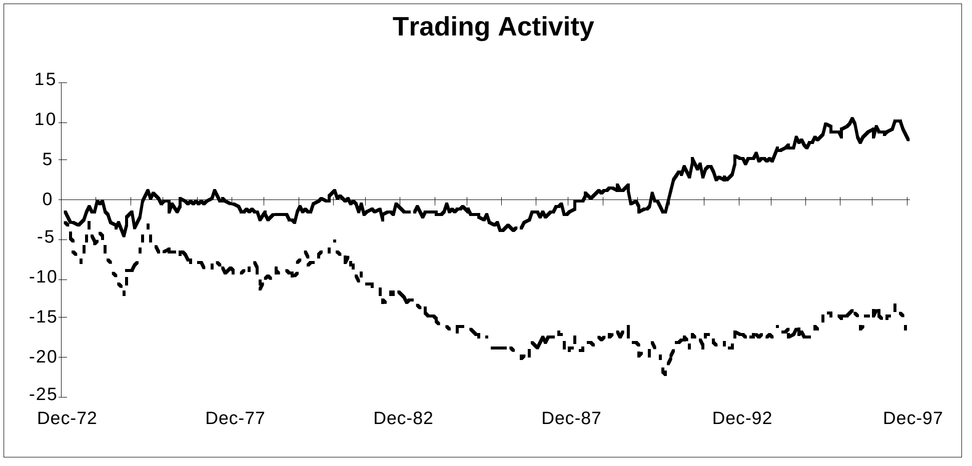

Dividend Yield
Number of estimation universe assets
900
800
700
600
500
400
300
200
100
0
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
Exposure (standard deviation)

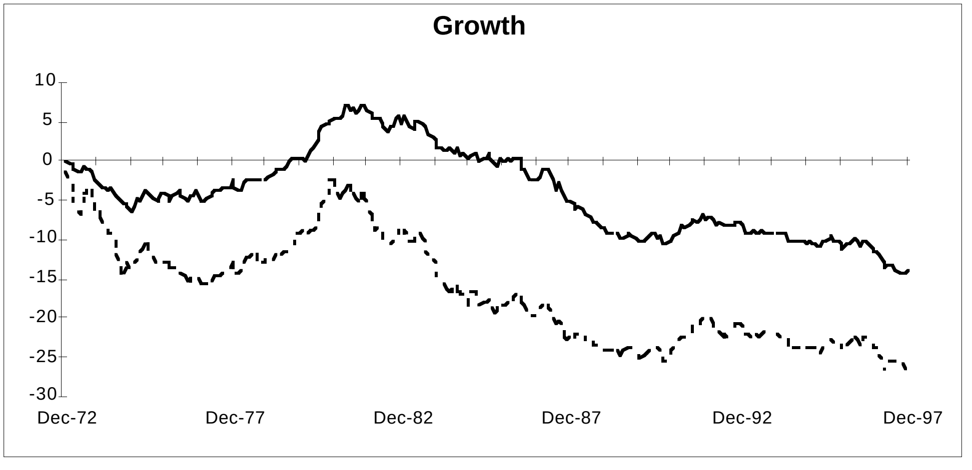
Appendix D: US-E3 Risk Index Factor Returns

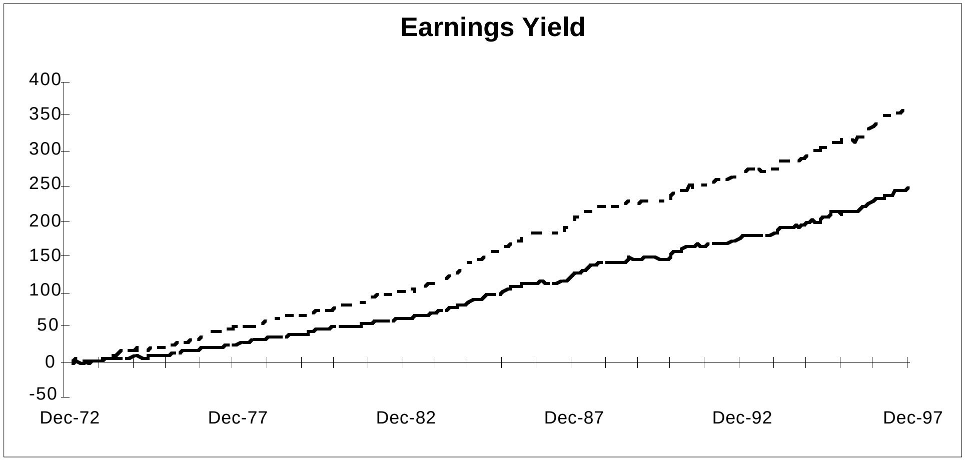
The following pages display factor return charts for US-E3's risk indices over the period December 1974–December 1997. The *single risk index return* is the return that results from a regression of asset return against a single risk index and all industries. The *multiple risk index return* is the return that results from a regression of asset return against all risk indices and all industries. The single risk index return is the return that results from a naive model; it is shown here for comparison purposes only. The multiple risk index return is the "true" factor return; it is the return used to explain performance and to construct the factor covariance matrix.

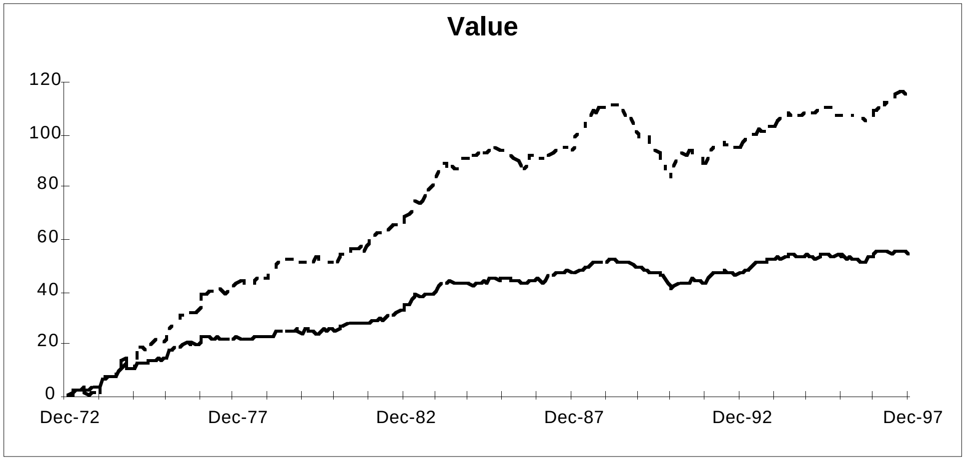

Volatility
30
20
10
0
-10
-20
-30
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97
Momentum
10
5
0
-5
-10
-15
-20
-25
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97
Size
10
0
-10
-20
-30
-40
-50
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97
Single Risk Index
Multiple Risk Index

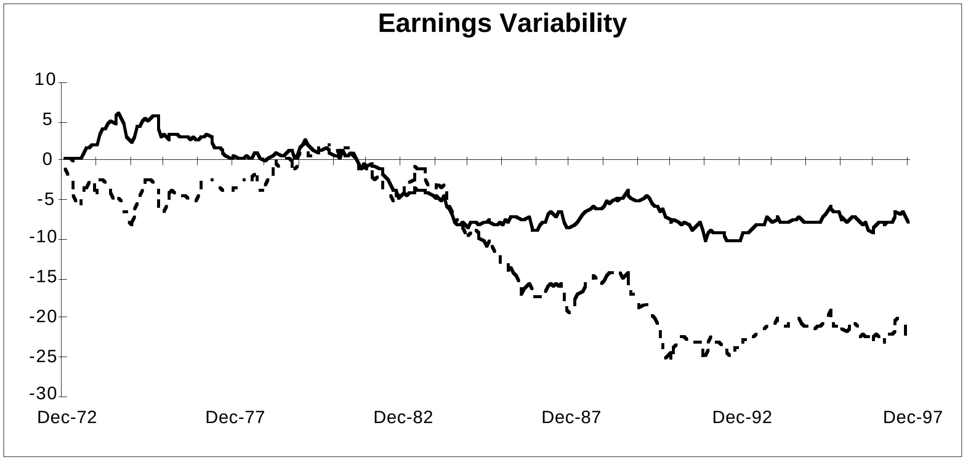

Size Nonlinearity
40
30
20
10
0
-10
-20
-30
-40
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97

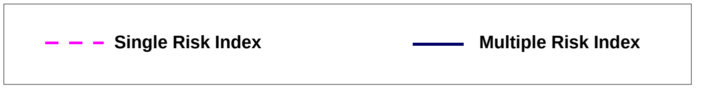

Trading Activity
15
10
5
0
-5
-10
-15
-20
-25
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97

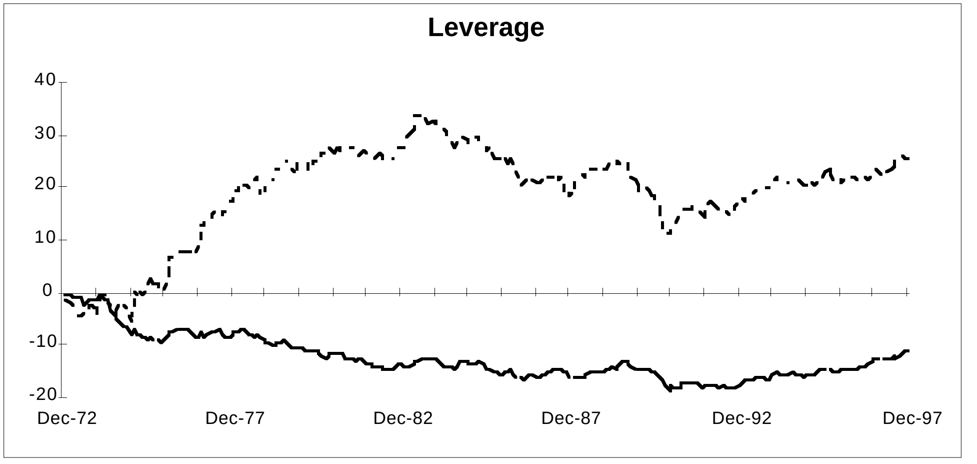

Growth
10
5
0
-5
-10
-15
-20
-25
-30
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97

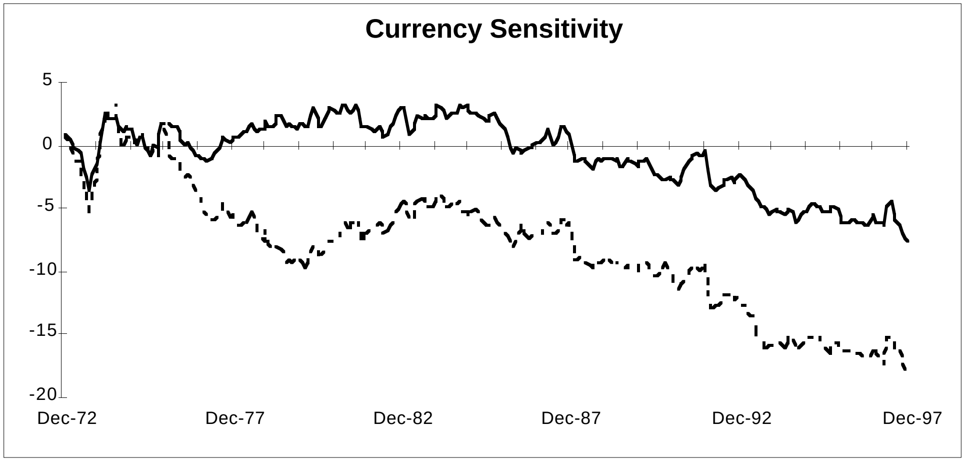

Earnings Yield
400
350
300
250
200
150
100
50
0
-50
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97

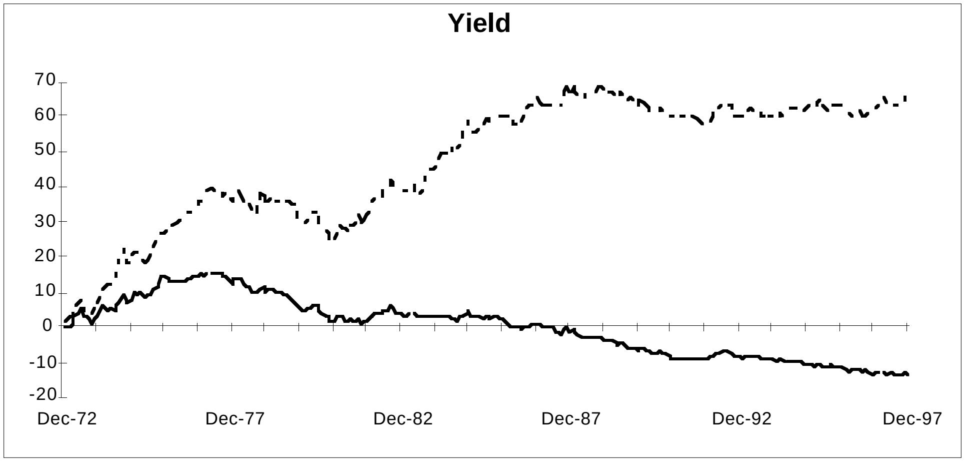

Value
120
100
80
60
40
20
0
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97


Earnings Variability
10
5
0
-5
-10
-15
-20
-25
-30
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97


Single Risk Index
Multiple Risk Index


Leverage
40
30
20
10
0
-10
-20
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97


Currency Sensitivity
5
0
-5
-10
-15
-20
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97


Yield
70
60
50
40
30
20
10
0
-10
-20
Dec-72
Dec-77
Dec-82
Dec-87
Dec-92
Dec-97

Glossary

active management	The pursuit of investment returns in excess of a specified benchmark.
active return	Return relative to a benchmark. If a portfolio's return is 5%, and the benchmark's return is 3%, then the portfolio's active return is 2%.
active risk	The risk (annualized standard deviation) of the active return. Also called tracking error.
alpha	The expected residual return. Beyond the pages of this book, alpha is sometimes defined as the expected exceptional return and sometimes as the realized residual or exceptional return.
arbitrage	To profit because a set of cash flows has different prices in different markets.
Arbitrage Pricing Theory (APT)	Developed in the late 1970s, the theory which asserts that securities and portfolio returns are based on the expected returns attributable to an unknown number of underlying factors. APT provides a complementary alternative to its precursor, the Capital Asset Pricing Model.
benchmark	A reference portfolio for active management. The goal of the active manager is to exceed the benchmark return.
beta	The sensitivity of a portfolio (or asset) to a benchmark. For every 1% return to the benchmark, we expect a $\beta \cdot 1\%$ return to the portfolio.
beta, historical	Historical measure of the response of a company's return to the market return, ordinarily computed as the slope coefficient in a 60-month historical regression.
beta, predicted	Predicted systematic risk coefficients (predictive of subsequent response to market return) that are derived, in whole or in part, from the fundamental operating characteristics of a company. Also called fundamental beta.

breadth	The number of independent forecasts available per year. A stock picker forecasting returns to 100 stocks every quarter exhibits a breadth of 400, assuming each forecast is independent (based on separate information).
Capital Asset Pricing Model (CAPM)	The simplest version states that the expected excess return on securities will be exactly in proportion to their systematic risk coefficient, or beta. The CAPM implies that total return on any security is equal to the risk-free return, plus the security's beta, multiplied by the expected market excess return.
certainty equivalent return	The certain (zero risk) return an investor would trade for a given (larger) return with an associated risk. For example, a particular investor might trade an expected 3% active return with 4% risk for a certain active return of 1.4%.
characteristic portfolio	A portfolio which efficiently represents a particular asset characteristic. For a given characteristic, it is the minimum risk portfolio with portfolio characteristic equal to 1. For example, the characteristic portfolio of asset betas is the benchmark. It is the minimum risk beta = 1 portfolio.
coefficient of determination (R^2)	*See* R-squared.
common factor	An element of return that influences many assets. According to multiple-factor risk models, the common factors determine correlations between asset returns. Common factors include industries and risk indices.
constraint	In portfolio optimization, a limitation imposed upon the portfolio so that it will have desired characteristics.
correlation	A statistical term giving the strength of linear relationship between two random variables. It is a pure number, ranging from -1 to +1: +1 indicates a perfect positive linear relationship; -1 indicates a perfect negative linear relationship; 0 indicates no linear relationship. For jointly distributed random variables, correlation is often used as a measure of strength of relationship, but it fails when a nonlinear relationship is present.

covariance	The tendency of different random investment returns to have similar outcomes, or to "covary." When two uncertain outcomes are positively related, covariance is positive, and conversely, negatively related outcomes have negative covariances. The magnitude of covariance measures the strength of the common movement. For the special case of a return's covariance with itself, the simplified name of variance is used. Covariance can be scaled to obtain the pure number, correlation, that measures the closeness of the relationship without its magnitude.
descriptor	A variable describing assets, used as an element of a risk index. For example, a volatility risk index, distinguishing high volatility assets from low volatility assets, could consist of several descriptors based on short-term volatility, long-term volatility, systematic and residual volatility, etc.
Dividend Discount Model (DDM)	A model of asset pricing based on discounting the future expected dividends.
dividend yield	The dividend per share divided by the price per share. Also known as the yield.
earnings yield	The earnings per share divided by the price per share.
efficient frontier	A set of portfolios, one for each level of expected return, with minimum risk. We sometimes distinguish different efficient frontiers based on additional constraints, e.g., the fully invested efficient frontier.
exceptional return	Residual return plus benchmark timing return. For a given asset with beta equal to 1, if its residual return is 2%, and the benchmark portfolio exceeds its consensus expected returns by 1%, then the asset's exceptional return is 3%.
excess return	Return relative to the risk-free return. If an asset's return is 3% and the risk-free return is 0.5%, then the asset's excess return is 2.5%.
factor portfolio	The minimum risk portfolio with unit exposure to the factor and zero exposures to all other factors. The excess return to the factor portfolio is the factor return.

factor return The return attributable to a particular common factor. We decompose asset returns into a common factor component, based on the asset's exposures to common factors times the factor returns, and a specific return.

information coefficient The correlation of forecast return with their subsequent realizations. A measure of skill.

information ratio The ratio of annualized expected residual return to residual risk. A central measurement for active management, value added is proportional to the square of the information ratio.

market The portfolio of all assets. We typically replace this abstract construct with a more concrete benchmark portfolio.

modern portfolio theory (MPT) The theory of portfolio optimization which accepts the risk/reward tradeoff for total portfolio return as the crucial criterion. Derived from Markowitz's pioneering application of statistical decision theory to portfolio problems, optimization techniques and related analysis are increasingly applied to investments.

multiple-factor model (MFM) A specification for the return process for securities. This model states that the rate of return on any security is equal to the weighted sum of the rates of return on a set of common factors, plus the specific return on the security, where the weights measure the exposures (or sensitivity) of the security to the factor. These exposures are identified with microeconomic characteristics, or descriptors of the firms (*see* descriptor).

Several simplifications of this model have been used historically. If there is only one factor, it becomes a *single-factor model*; if this one factor is identified with an index, it is called a *single-index model*; if the single-factor is identified with the market factor, it becomes the *market model*. Depending on the statistical specification, some of these could become a *diagonal model*, which simply indicates that the covariance matrix between security returns is (or can easily be transformed into) a diagonal matrix.

normal A benchmark portfolio.

normalization The process of transforming a random variable into another form with more desirable properties. One example is standardization in which a constant (usually the mean) is subtracted from each number to shift all numbers uniformly. Then each number is divided by another constant (usually the standard deviation) to shift the variance.

optimization The best solution among all the solutions available for consideration. Constraints on the investment problem limit the region of solutions that are considered, and the objective function for the problem, by capturing the investor's goals correctly, providing a criterion for comparing solutions to find the better ones. The optimal solution is that solution among those admissible for consideration which has the highest value of the objective function. The first-order conditions for optimality express the tradeoffs between alternative portfolio characteristics to provide the optimum solution.

outlier A data observation that is very different from other observations. It is often the result of an extremely rare event or a data error.

passive management Managing a portfolio to match (not exceed) the return of a benchmark.

payout ratio The ratio of dividends to earnings. The fraction of earnings paid out as dividends.

performance analysis Evaluation of performance in relation to a standard or benchmark with the purpose of assessing manager skill.

performance attribution The process of attributing portfolio returns to causes. Among the causes are the normal position for the portfolio, as established by the owner of funds or the manager, as well as various active strategies, including market timing, common factor exposure, and asset selection. Performance attribution serves an ancillary function to the prediction of future performance, in as much as it decomposes past performance into separate components that can be analyzed and compared with the claims of the manager.

R-squared A statistic usually associated with regression analysis, where it describes the fraction of observed variation in data captured by the model. It varies between 0 and 1.

regression	A data analysis technique that optimally fits a model based on the squared differences between data points and model fitted points. Typically, regression chooses model coefficients to minimize the (possibly weighted) sum of these squared differences.
residual return	Return independent of the benchmark. The residual return is the return relative to beta times the benchmark return. To be exact, an asset's residual return equals its excess return minus beta times the benchmark excess return.
residual risk	The risk (annualized standard deviation) of the residual return.
risk	The uncertainty of investment outcomes. Technically, risk defines all uncertainty about the mean outcome, including both upside and downside possibilities. The more intuitive concept for risk measurement is the standard deviation of the distribution, a natural measure of spread. Variance, the square of the standard deviation, is used to compare independent elements of risk.
risk-free return	The return achievable with absolute certainty. In the U.S. market, short maturity Treasury bills exhibit effectively risk-free returns. The risk-free return is sometimes called the time premium, as distinct from the risk premium.
risk index	A common factor typically defined by some continuous measure, as opposed to a common industry membership factor defined as 0 or 1. Risk index factors include Volatility, Momentum, Size, and Value.
risk premium	The expected excess return to the benchmark.
score	A normalized asset return forecast. An average score is 0, with roughly two-thirds of the scores between -1 and 1. Only one-sixth of the scores lie above 1.
security market line	The linear relationship between asset returns and betas posited by the Capital Asset Pricing Model.
Sharpe ratio	The ratio of annualized excess returns to total risk.

significance (statistical significance)

A statistical term which measures the spread or variability of a probability distribution. The standard deviation is the square root of variance. Its intuitive meaning is best seen in a simple, symmetrical distribution, such as the normal distribution, where approximately two-thirds of all outcomes fall within ± 1 standard deviation of the mean, approximately 95 percent of all outcomes fall within ± 2 standard deviations, and approximately 99 percent of all outcomes fall within ± 2.5 standard deviations. The standard deviation of return—or, more properly, of the logarithm of return, which is approximately symmetrically distributed—is very widely used as a measure of risk for portfolio investments.

skill

The ability to accurately forecast returns. We measure skill using the information coefficient.

specific return

The part of the excess return not explained by common factors. The specific return is independent of (uncorrelated with) the common factors and the specific returns to other assets. It is also called the **idiosyncratic return**.

specific risk

The risk (annualized standard deviation) of the specific return.

standard error

The standard deviation of the error in an estimate. A measure of the statistical confidence in the estimate.

standardization

Standardization involves setting the zero point and scale of measurement for a variable. An example might be taken from temperature, where the centigrade scale is standardized by setting zero at the freezing point of water and establishing the scale (the centigrade degree) so that there are 100 units between the freezing point of water and the boiling point of water. Standardization for risk indices and descriptors in BARRA equity models sets the zero value at the capitalization-weighted mean of the companies in the universe and sets the unit scale equal to one cross-sectional standard deviation of that variable among the estimation universe.

systematic return

The part of the return dependent on the benchmark return. We can break excess returns into two components: systematic and residual. The systematic return is the beta times the benchmark excess return.

systematic risk

The risk (annualized standard deviation) of the systematic return.

t-statistic

The ratio of an estimate to its standard error. The *t*-statistic can help test the hypothesis that the estimate differs from zero. With some standard statistical assumptions, the probability that a variable with a true value of zero would exhibit a *t*-statistic greater than 2 in magnitude is less than 5%.

tracking error

See active risk.

transaction costs

The costs incurred for a portfolio when securities are changed for other securities. Transaction costs are deducted from the value of the portfolio directly, rather than paid as fees to the money manager. These costs arise from three sources: (1) commissions and taxes paid directly in cash; (2) the typical "dealer's spread" (or one-half of this amount) earned by a dealer, if any, who acts as an intermediary between buyer and seller; and (3) the net advantage or disadvantage earned by giving or receiving accommodation to the person on the other side of the trade. The third component averages out to zero across all trades, but it may be positive or negative, depending on the extent to which a trader, acting urgently, moves the market against the selected strategy.

universe

The list of all assets eligible for consideration for inclusion in a portfolio. At any time, some assets in the universe may be temporarily ruled out because they are currently viewed as overvalued. However, the universe should contain all securities that might be considered for inclusion in the near term if their prices move to such an extent that they become undervalued. Universe also defines the normal position of a money manager, equating the normal holding with the capitalization-weighted average of the securities in the universe or followed list.

utility

A measure of the overall desirability or goodness of a person's situation. In the theory of finance, utility is the desirability of a risky series of outcomes. The utility (or expected utility) of a set of risky outcomes is assumed to measure its goodness, so that a package with higher utility is always preferred to one with lower utility. In portfolio theory, utility is almost always defined by a function of the mean and variance of portfolio outcomes, which is then called a mean/variance utility function. The further assumption that the utility function is linear in its two arguments (mean and variance) results in a linear mean/variance utility function (LMVU).

value added The utility, or risk-adjusted return, generated by an investment strategy: the return minus a risk aversion constant times the variance. The value added depends on the performance of the manager and the preferences of the owner of the funds.

variance A statistical term for the variability of a random variable about its mean. The variance is defined as the expected squared deviation of the random variable from its mean—that is, the average squared distance between the mean value and the actually observed value of the random variable. When a portfolio includes several independent elements of risk, the variance of the total arises as a summation of the variances of the separate components.

volatility A loosely-defined term for risk. Here we define volatility as the annualized standard deviation of return.

yield *See* dividend yield.

Index

D

E

F

G

H

V

W

Y

Contributors

Editors

Ronald N. Kahn
Pamela Brougham
Paul Green

Consultants

Ken Hui
Tom Leimkuhler
Scott Scheffler
Aamir Sheikh

Production Services

Susan McIntosh