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 interestrate 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 interestrate 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, 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 1-1

Movements in the dollar.

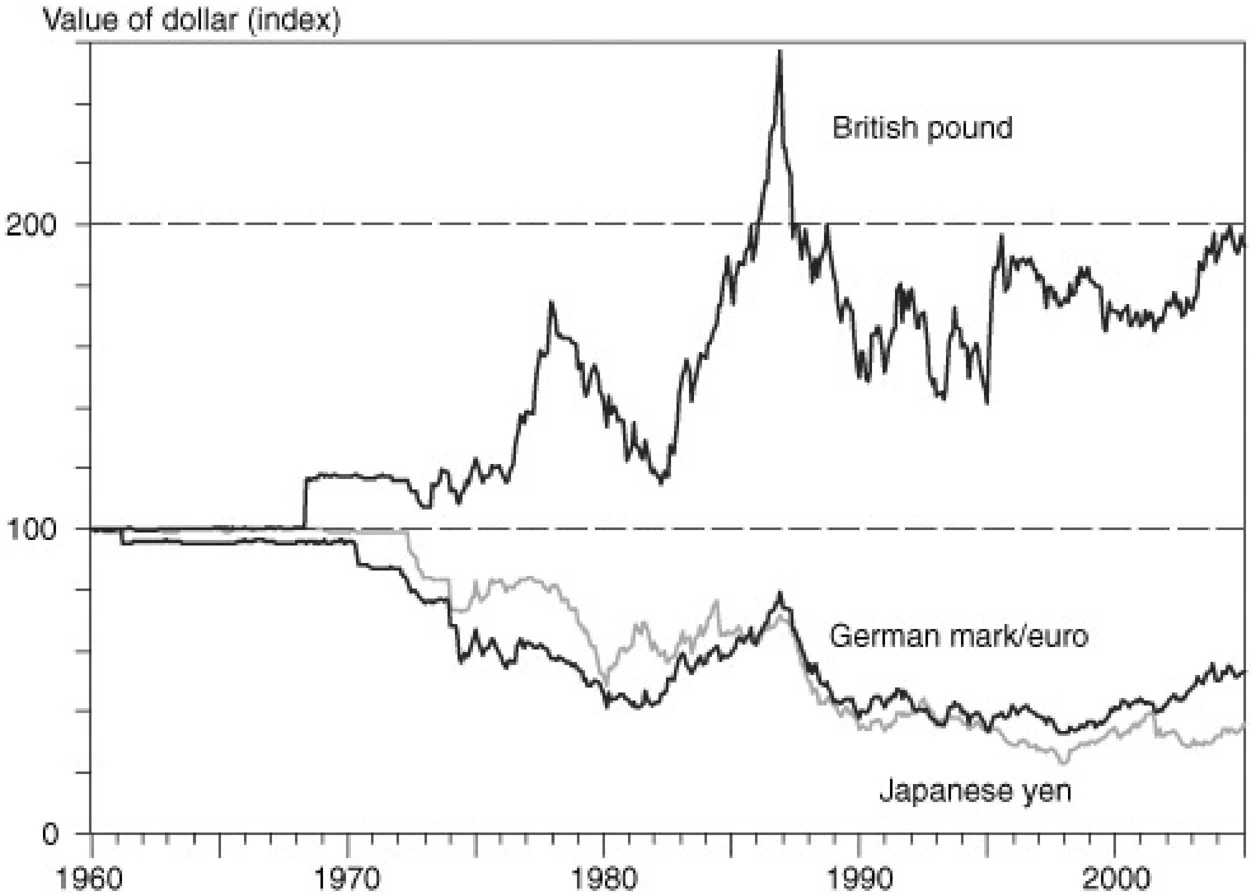


Figure 1-3 shows that oil prices also have fluctuated widely. The sharp oil price increases of the 1970s seem correlated with increases in bond yields. These oil shocks also had an impact on national stock markets, which are displayed in Figure 1-4. Indeed, the great bear market of 1974–1975 was a global occurrence triggered by a threefold increase in the price of crude oil. This episode shows that it is difficult to understand financial risk without a good grasp of the underlying economics, as well as the links between major risk categories.

FIGURE 1-2

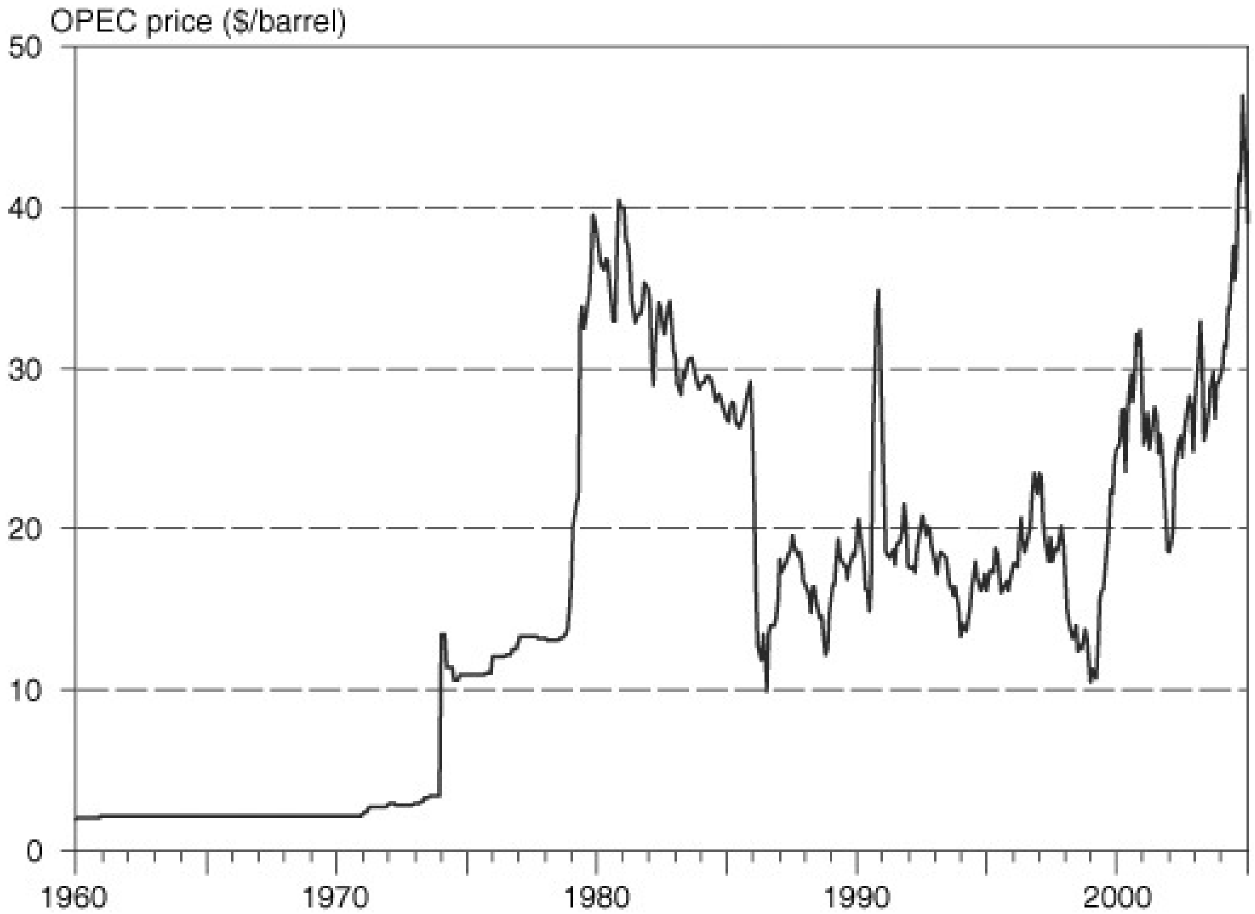
Movements in U.S. interest rates.

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 interestrate 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 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.

FIGURE 1-4 Movements in stock prices.

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.

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

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

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

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.

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.

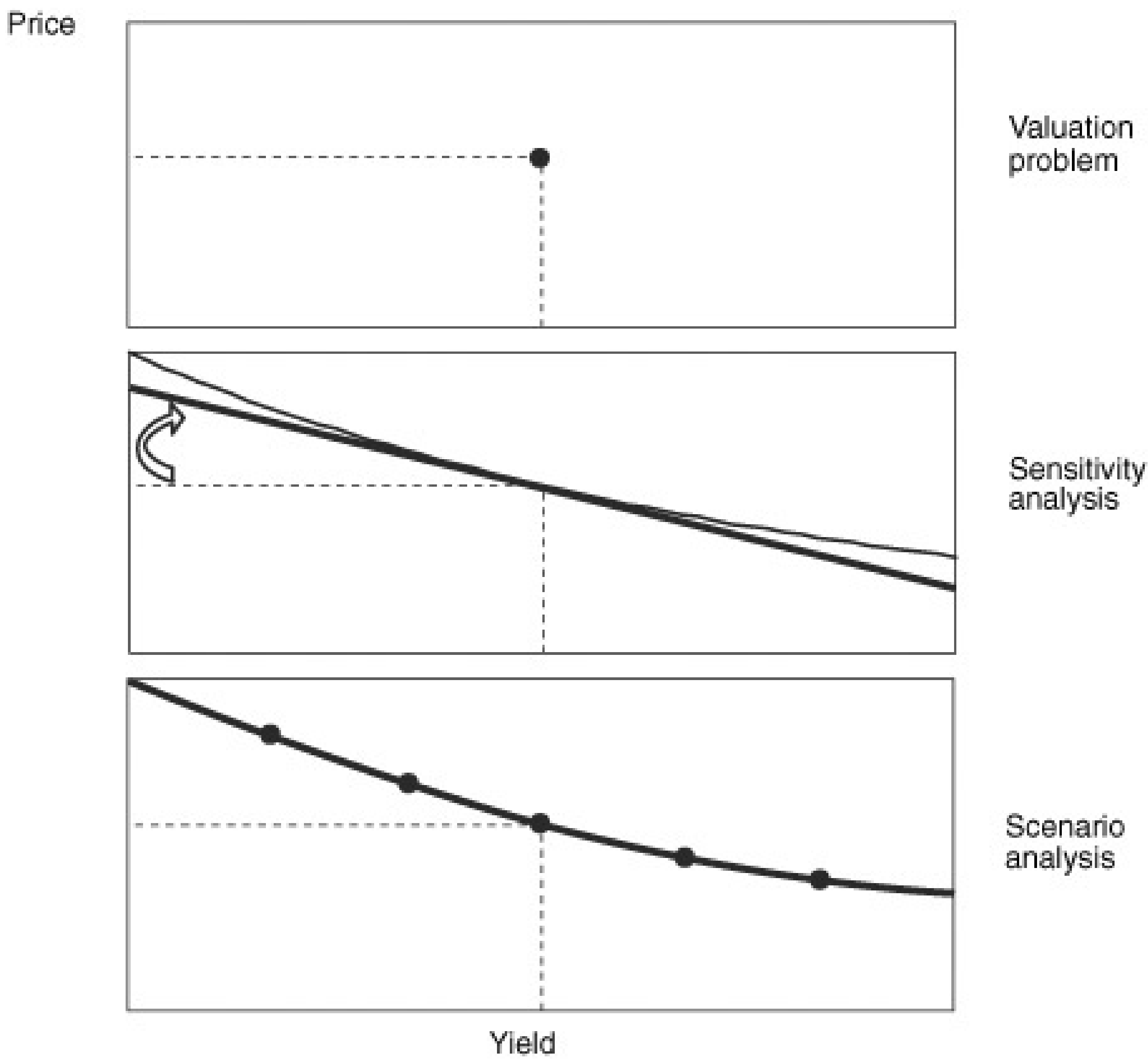
Instead, the risk manager needs to use ex ante, or forward-looking risk controls. A limit could be placed on the *notional amount*. This is not sufficient, however. For the same notional amount, some bonds have extreme risks and others no risk. The risk manager needs to know how the instruments respond 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 interestrate 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 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.

FIGURE 1-5

Conventional risk-measurement methods.

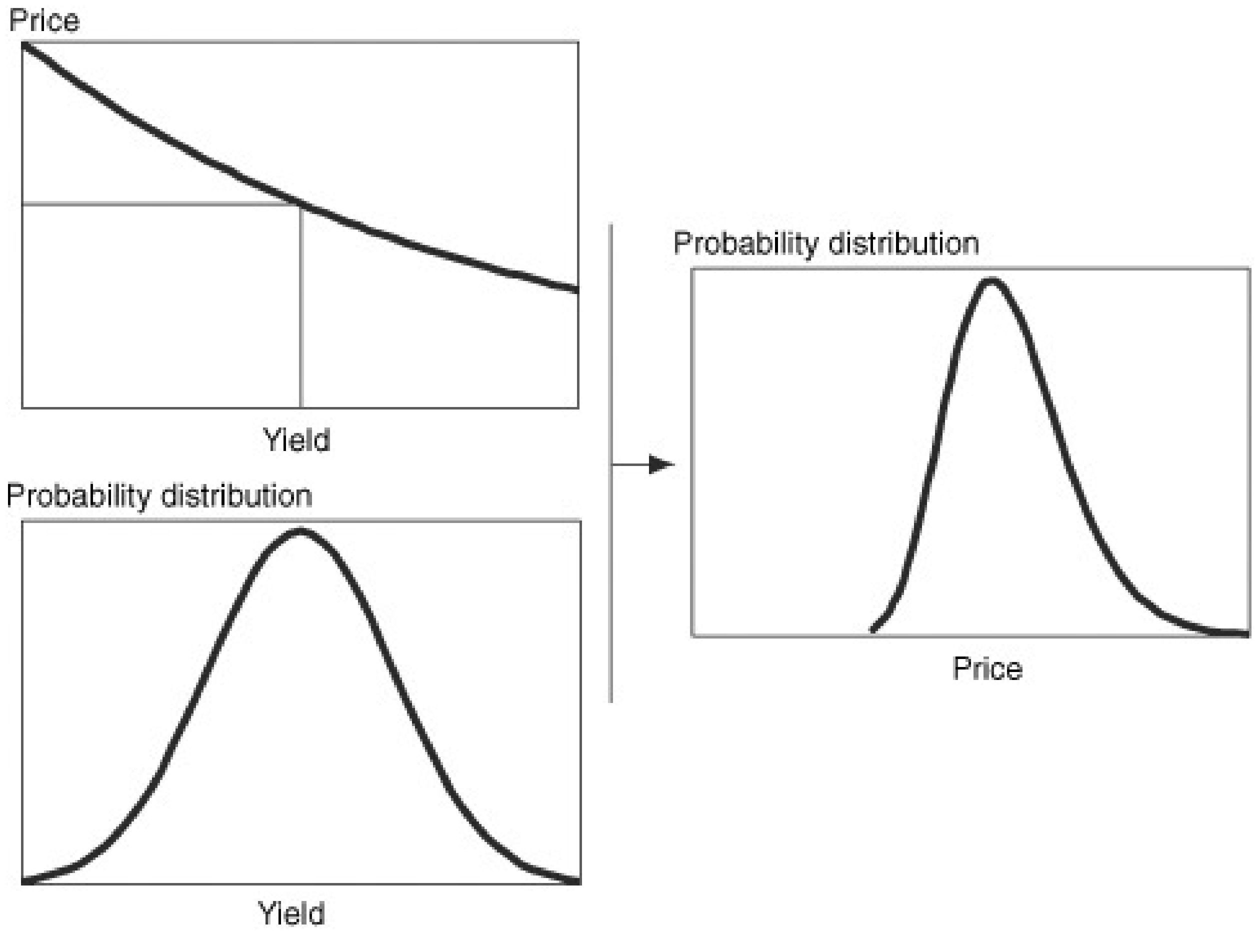


This approach is still incomplete, however. It does not consider the volatility of the risk factors, which could vary across markets, nor their correlations. It cannot be used to set consistent limits across bonds and equities, for example. Because sensitivity measures are not additive, they cannot be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

BOX 1-1

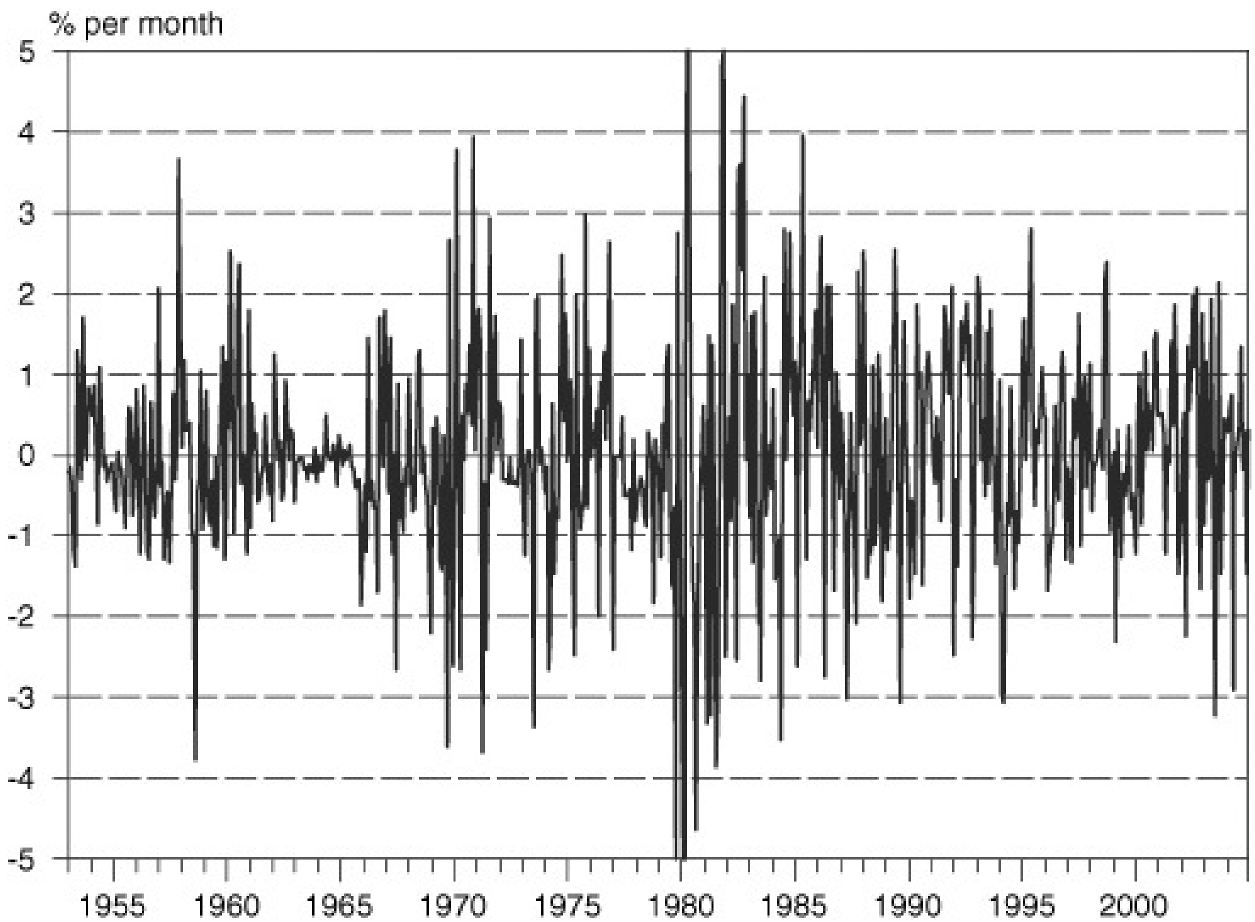
THE ORIGINS OF VAR

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.



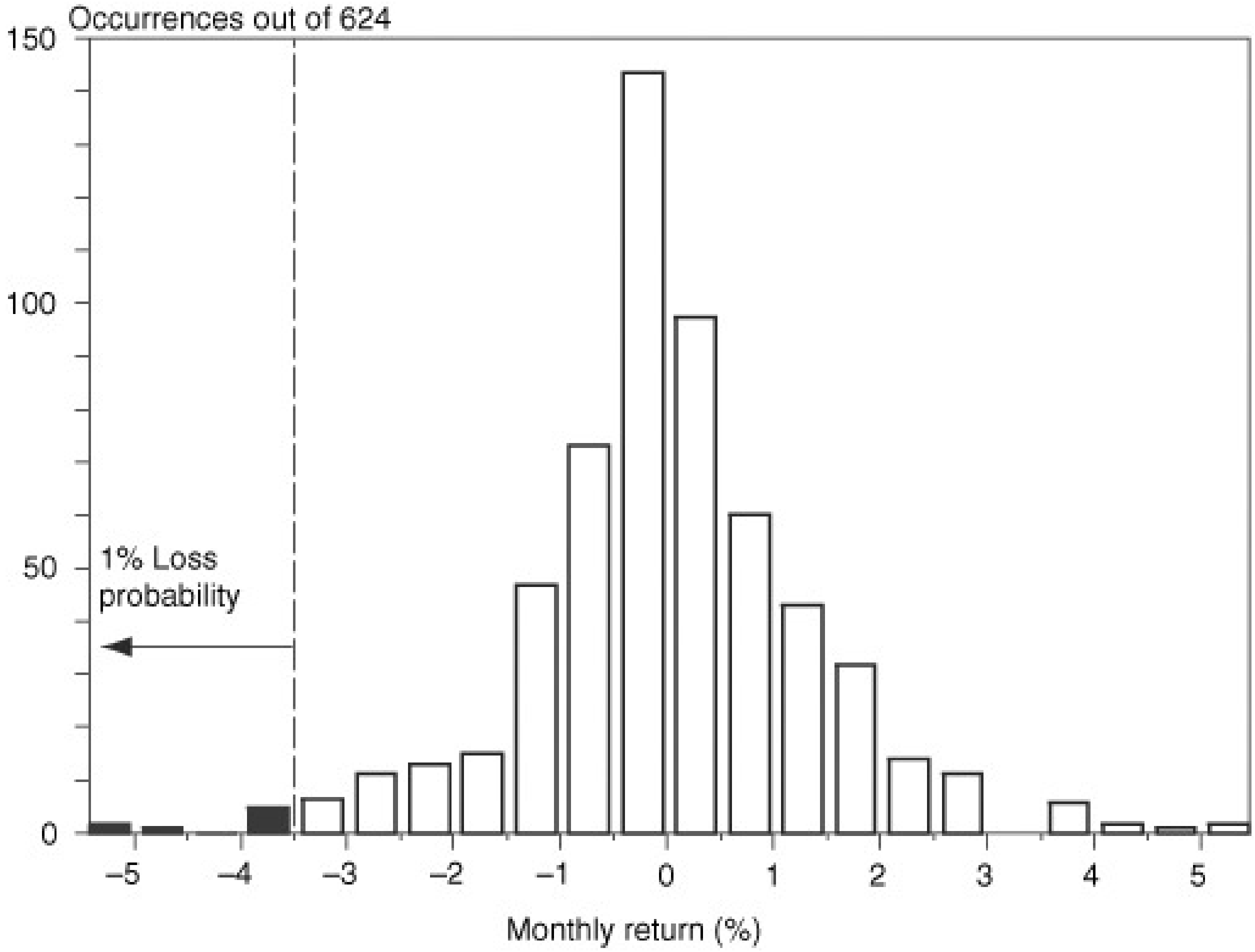
Next, associate with each return a probability of observing a lower value.

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 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.

FIGURE 1-8 Measuring value at risk.



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

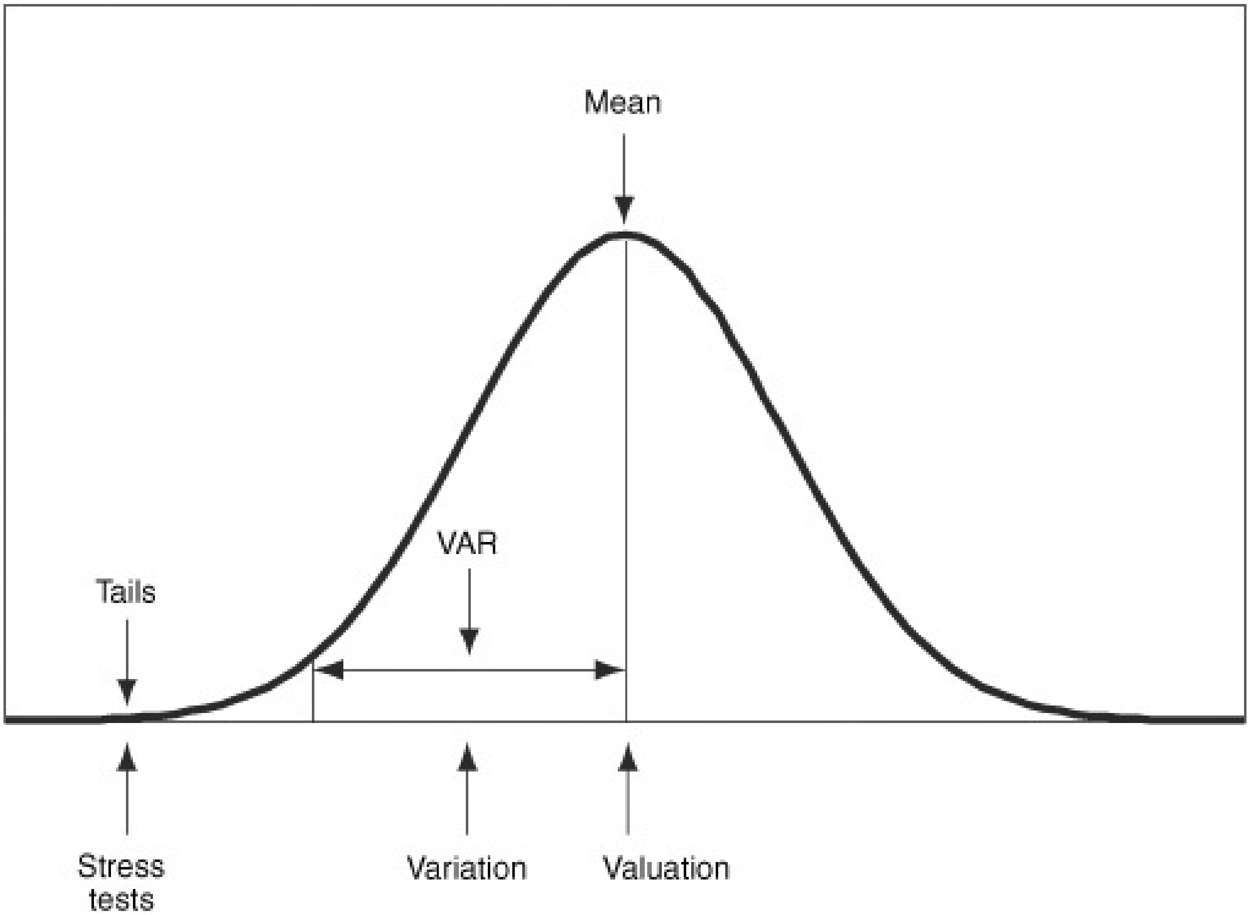
In a broad sense, VAR extends current valuation methods for derivatives instruments. To price options, for instance, we need to make an assumption about the distribution of the driving risk factor. The option then is priced by taking the present value of the *expected* option value at maturity. This 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 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]

FIGURE 1-9
Different views of the payoff distribution.



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.

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 interestrate 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 interestrate 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 interestrate swaps entered by city councils in Britain. The municipalities of Hammersmith and Fulham had taken large positions in interestrate 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 not, governments themselves contribute to the risks through poor bank supervision, ill-advised government intervention, or unsustainable economic policies.[1]

TABLE 2-2
Cost of Financial Insolvencies

Country	Scope	Cost	
		% 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.

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 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 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 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.

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.

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 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:

> 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 inhouse 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 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:

- *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

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

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

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.

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:

$$\text{Capital} > \text{TRC} = \text{CRC} + \text{MRC} + \text{ORC} \quad (3.2)$$

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 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.

FIGURE 3-1

Application of Basel II Accord.

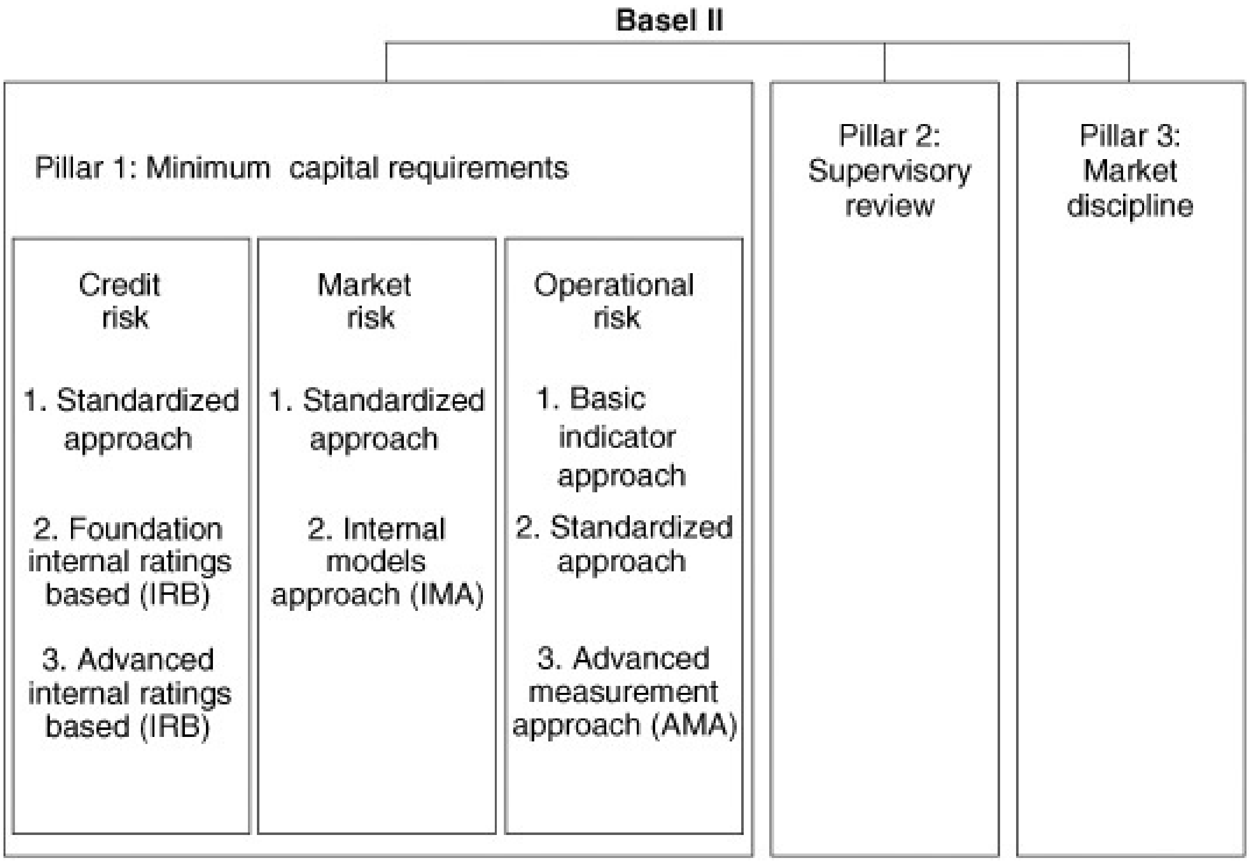


TABLE 3-2
New Basel Risk Weights: Standardized Approach

	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.

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 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:

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

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 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:

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.

- *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 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.

$$MRC_t^{IMA} = \max(k \frac{1}{60} \sum_{i=1}^{60} VAR_{t-i}, \quad VAR_{t-1}) + SRC_t \qquad (3.4)$$

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

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.

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

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).

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.

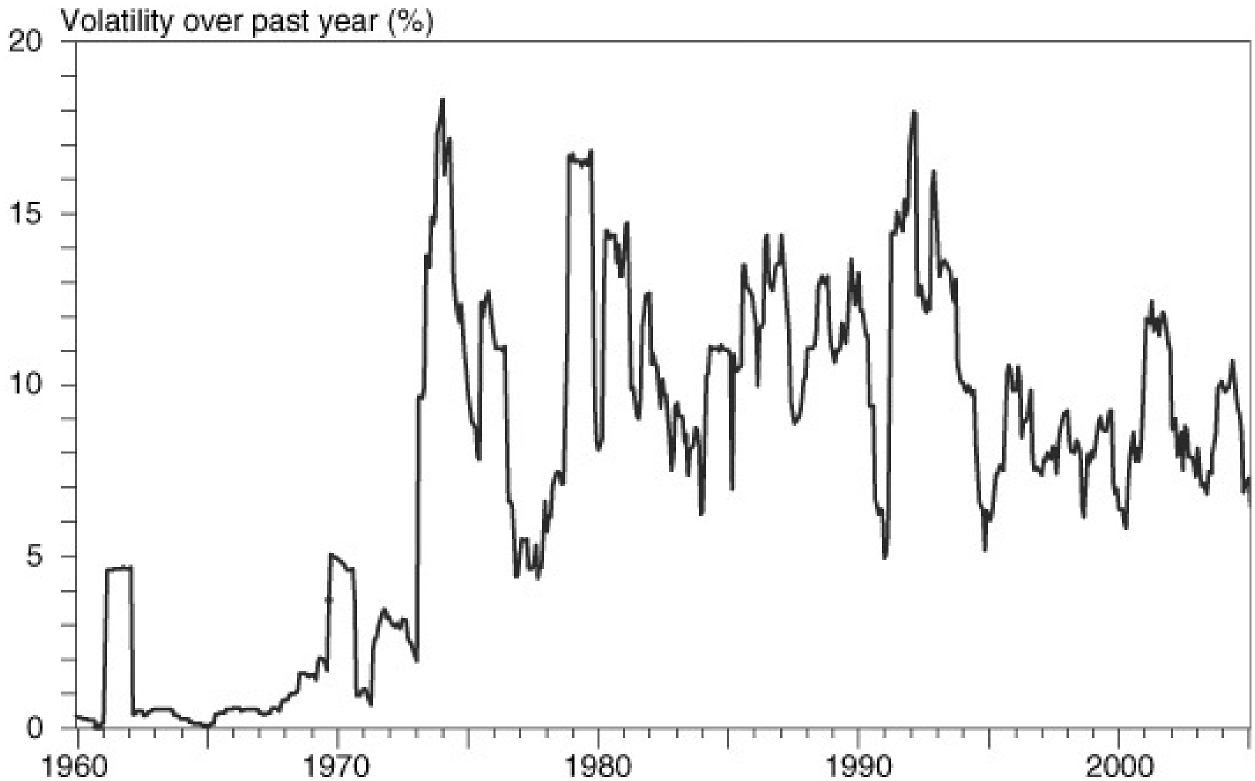


FIGURE 4-2
Volatility in interest rates.

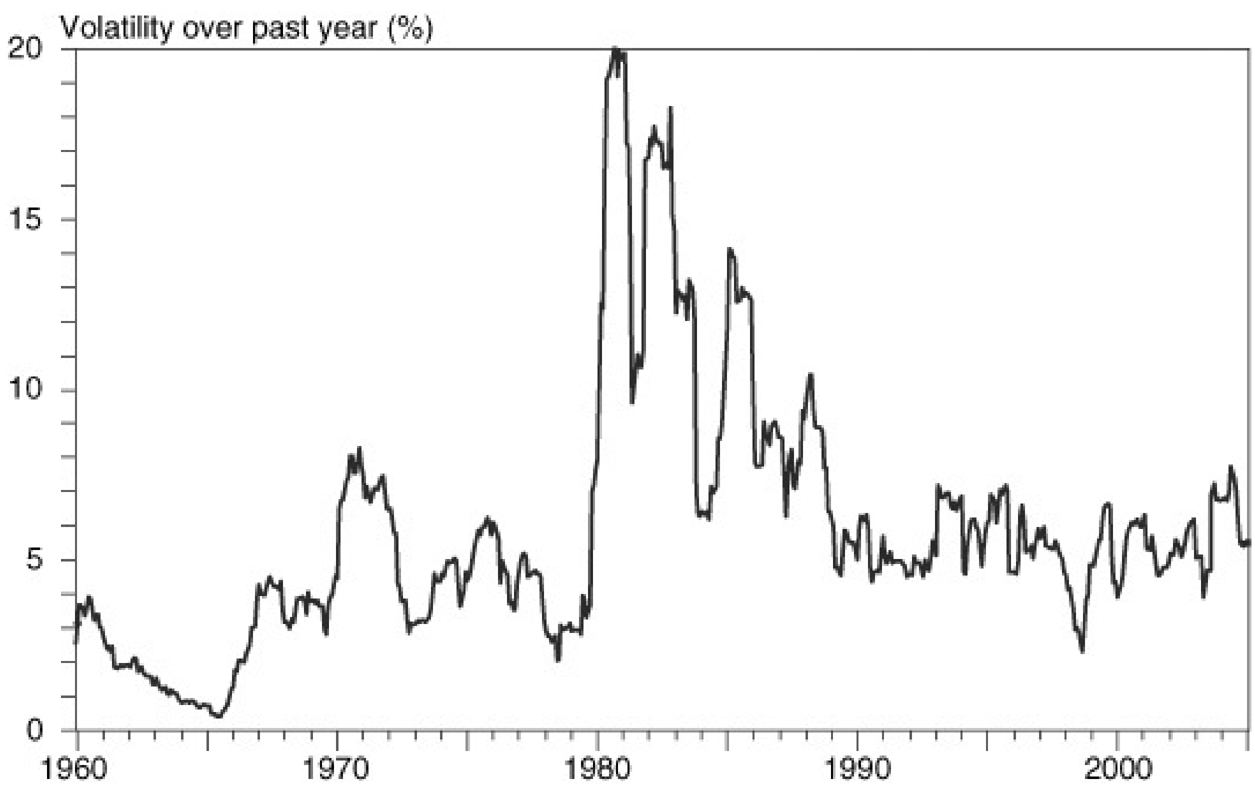


FIGURE 4-3
Volatility in oil prices.

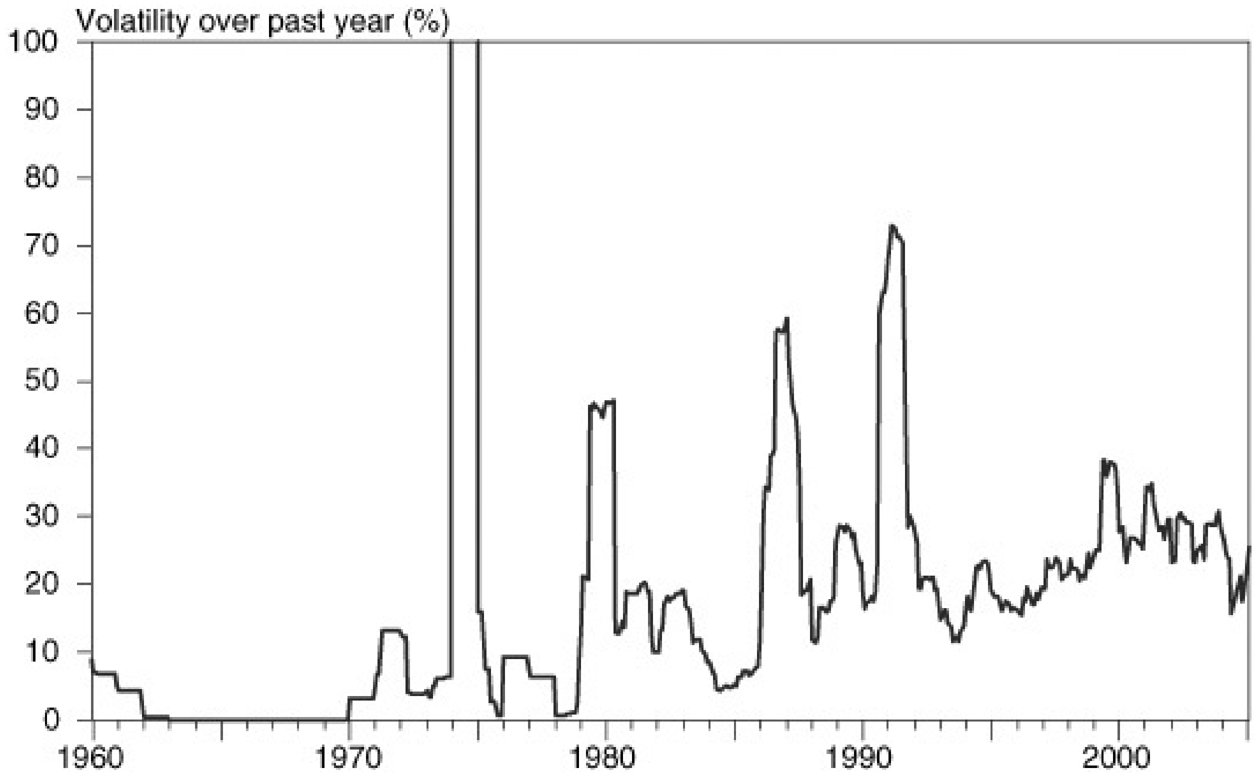
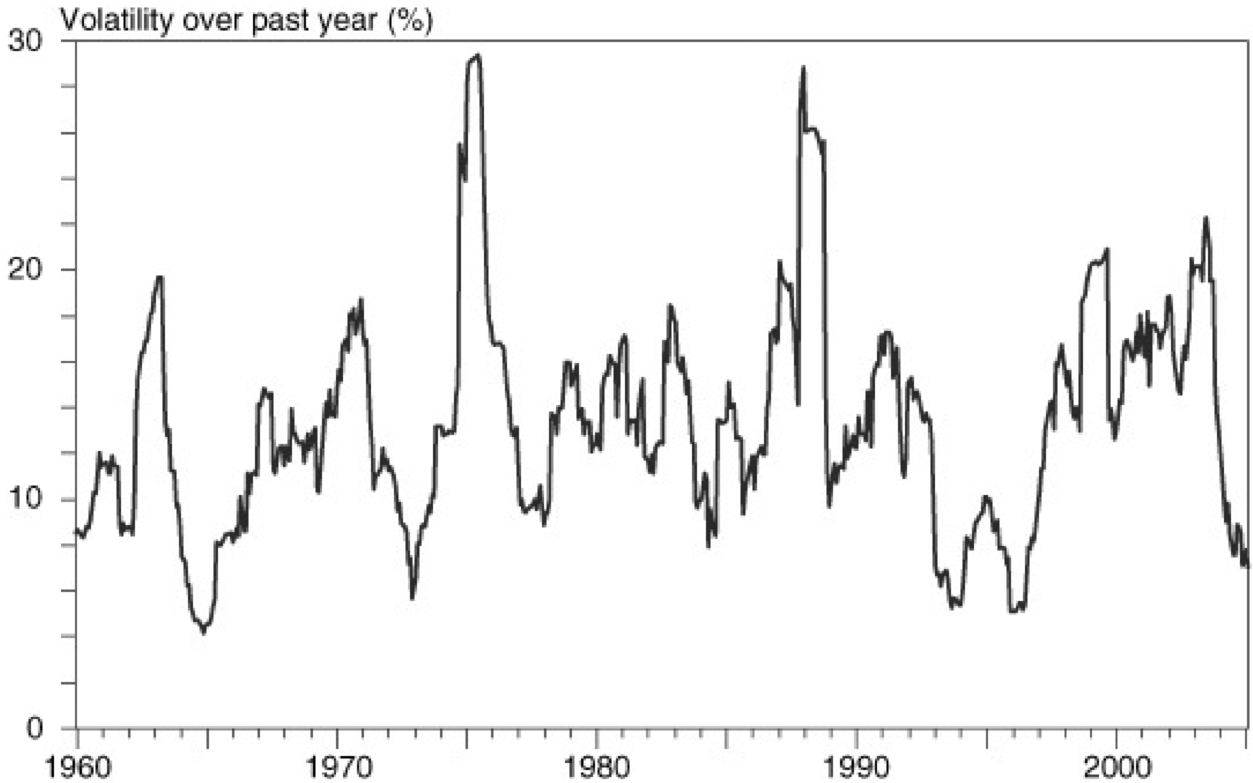


FIGURE 4-4
Volatility in stock prices.



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.

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:

$$\sum_{i=1}^{11} p_i = 1 \quad (4.1)$$

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

FIGURE 4-5
Distribution of payoff.

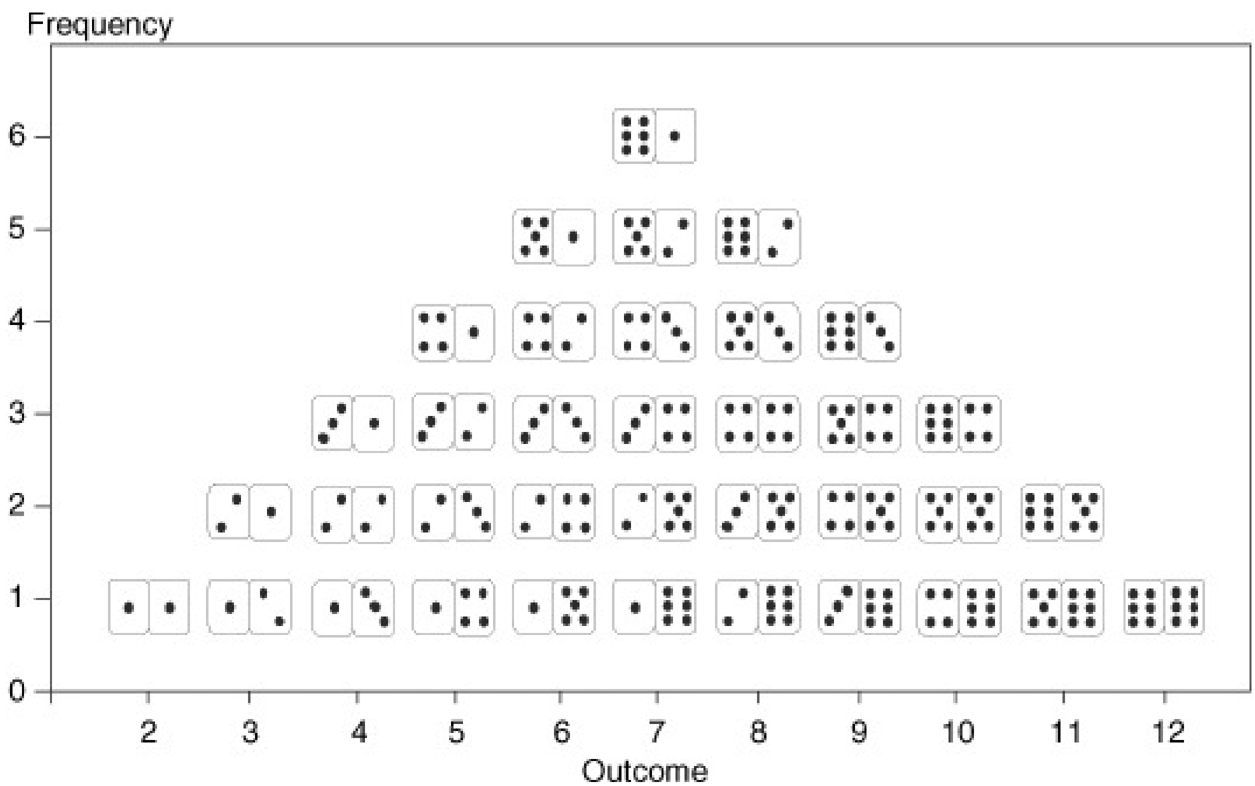


TABLE 4-1
Computing Expected Value and Standard Deviation

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:

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

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]

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

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:

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

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:

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

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), by Equation (4.5), and

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

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

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 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.

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

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.

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

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

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.

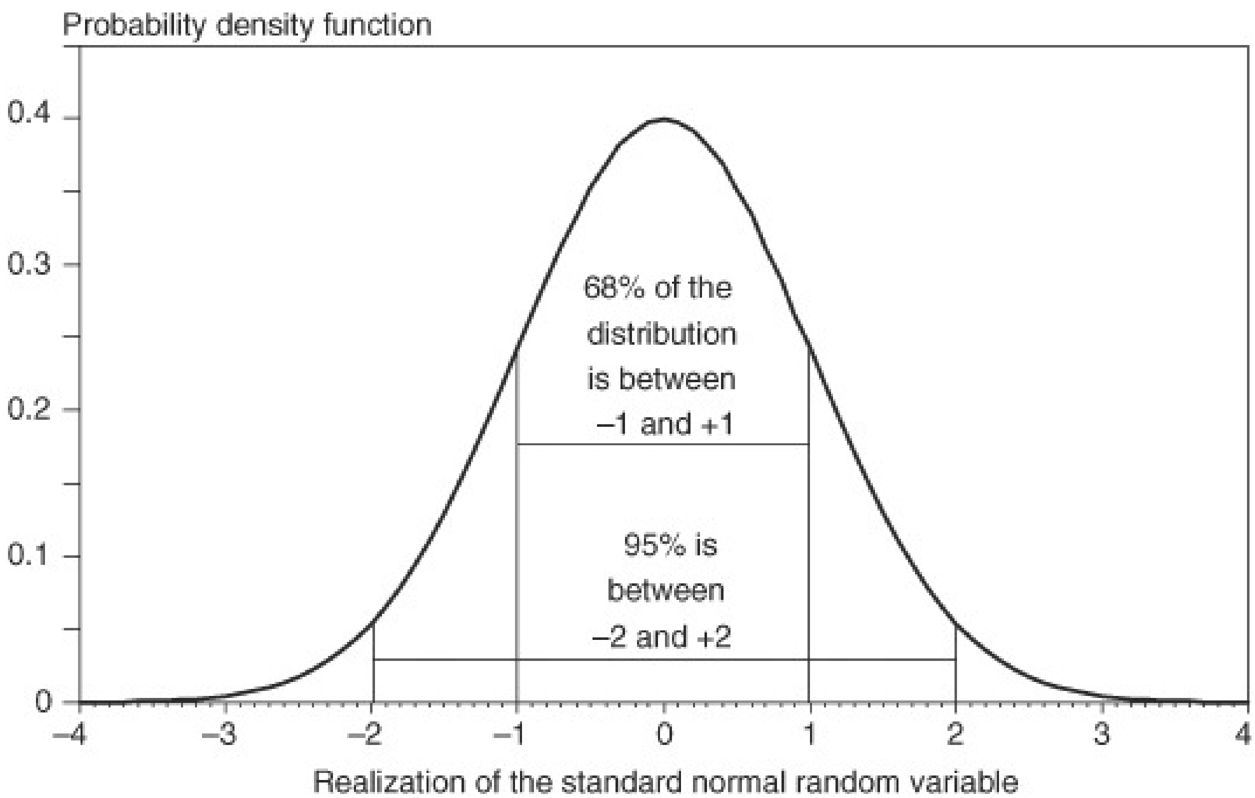
In practice, the normal density function is tabulated for a variable with mean zero and variance unity, called a *standard* normal variable ϵ. This then is transformed into a variable with any mean or standard deviation. Define X as

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

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

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

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

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 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 $\delta = 3 + 6/(n - 4)$.

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

Another useful distribution is the *generalized error distribution* (GED). Its pdf is where ν is a shape-defining parameter. This function is convenient because it includes the normal pdf as a special case, with $\nu = 2$. The pdf has fatter tails for values of ν lower than 2. Here, the scaling parameter λ ensures that the variance of *X* is unity.

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

FIGURE 4-7

Comparison of parametric distributions.

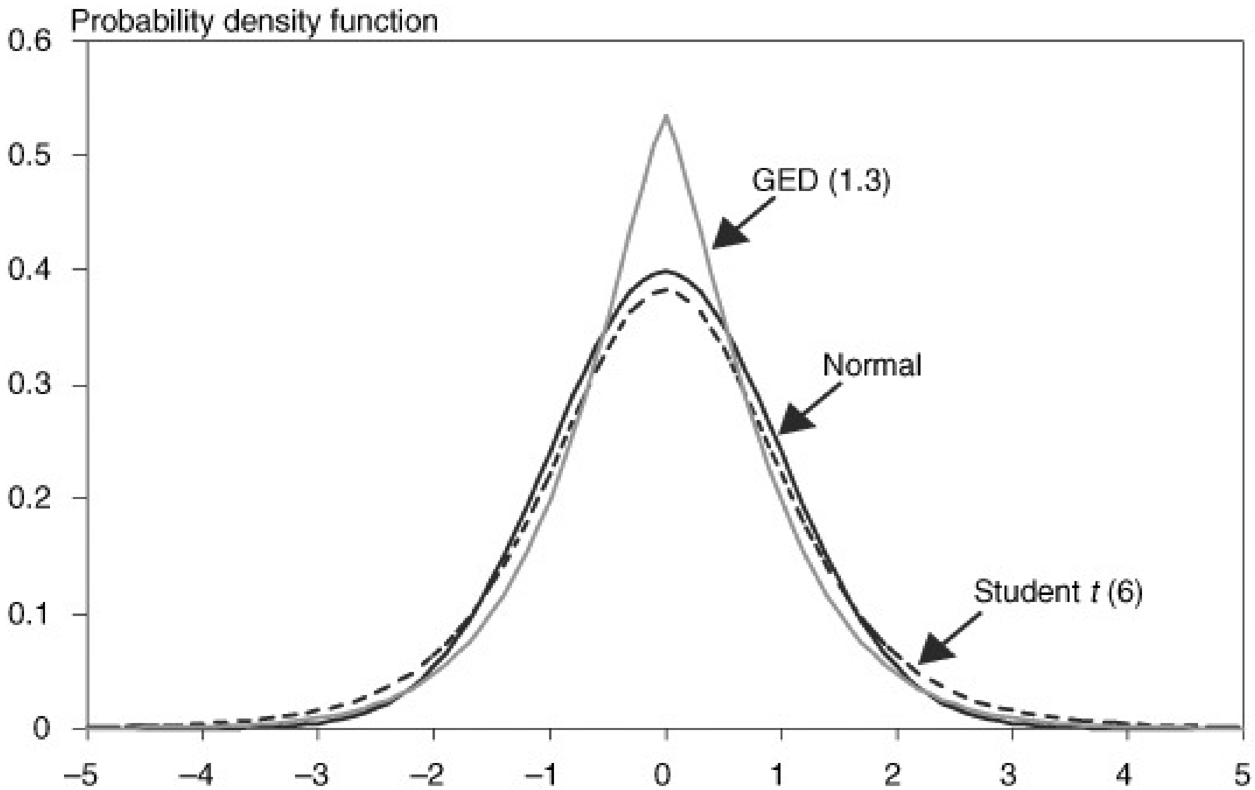


Figure 4-7 compares the normal distribution with a student *t* distribution with *n* = 6 and a GED distribution with ν = 1.3. These parameters typically describe financial data. We can see that both distributions have fatter tails than the normal density. This may be particularly important when assessing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

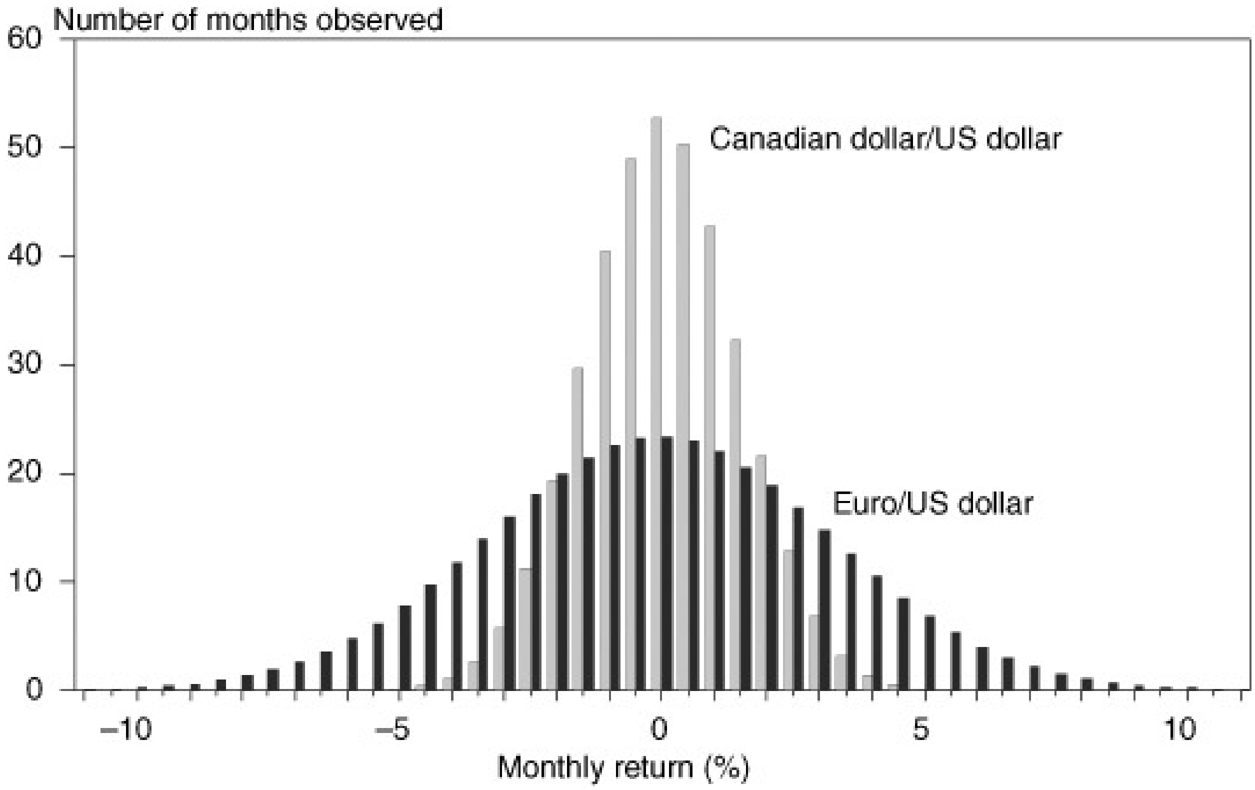
4.3 RISK

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) 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.

FIGURE 4-8

Comparison of currency distributions.



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 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.

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

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, . . ., 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,

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

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(EUR/BP)] = ln[S(EUR/$)] + ln[$S$($/BP)] = ln[S($/BP)] − ln[$S$($/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

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

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 $R_t = r_t - r^2 /$

$+ r_t^3 \ldots$, 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: and the variance, or second moment, $\sigma^2 = E[(X - \mu)^2]$ can be estimated by the sample variance:

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

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

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

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

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

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

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

$$\mathrm{SE}(\hat{\gamma}) = \sqrt{\frac{6}{T}} \qquad (4.34)$$

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

$$\mathrm{SE}(\hat{\delta}) = \sqrt{\frac{24}{T}} \qquad (4.35)$$

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,

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

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

$$\mu = \mu_{\text{annual}} T \qquad (4.37)$$

$$\sigma = \sigma_{\text{annual}} \sqrt{T} \qquad (4.38)$$

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 × 12 = −1.8 percent per annum. The volatility is 3.28 percent per month × $\sqrt{12}$ = 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 which is higher than in the i.i.d. case if $\rho > 0$. In general, the variance over N periods can be written as

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

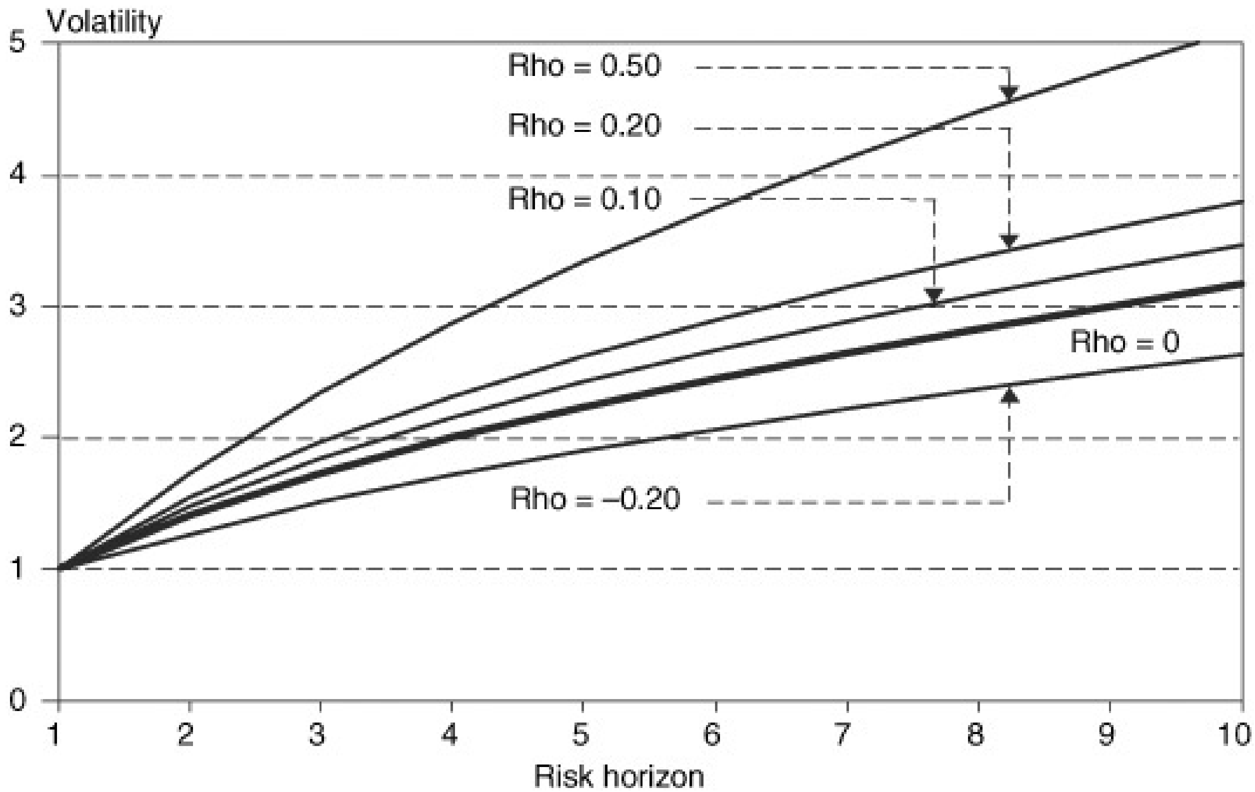
$$V\left(\sum_{i=1}^{N} X_{t+i}\right) = \sigma^2[N + 2(N-1)\rho + 2(N-2)\rho^2 + \cdots + 2(1)\rho^{N-1}] \tag{4.41}$$

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).

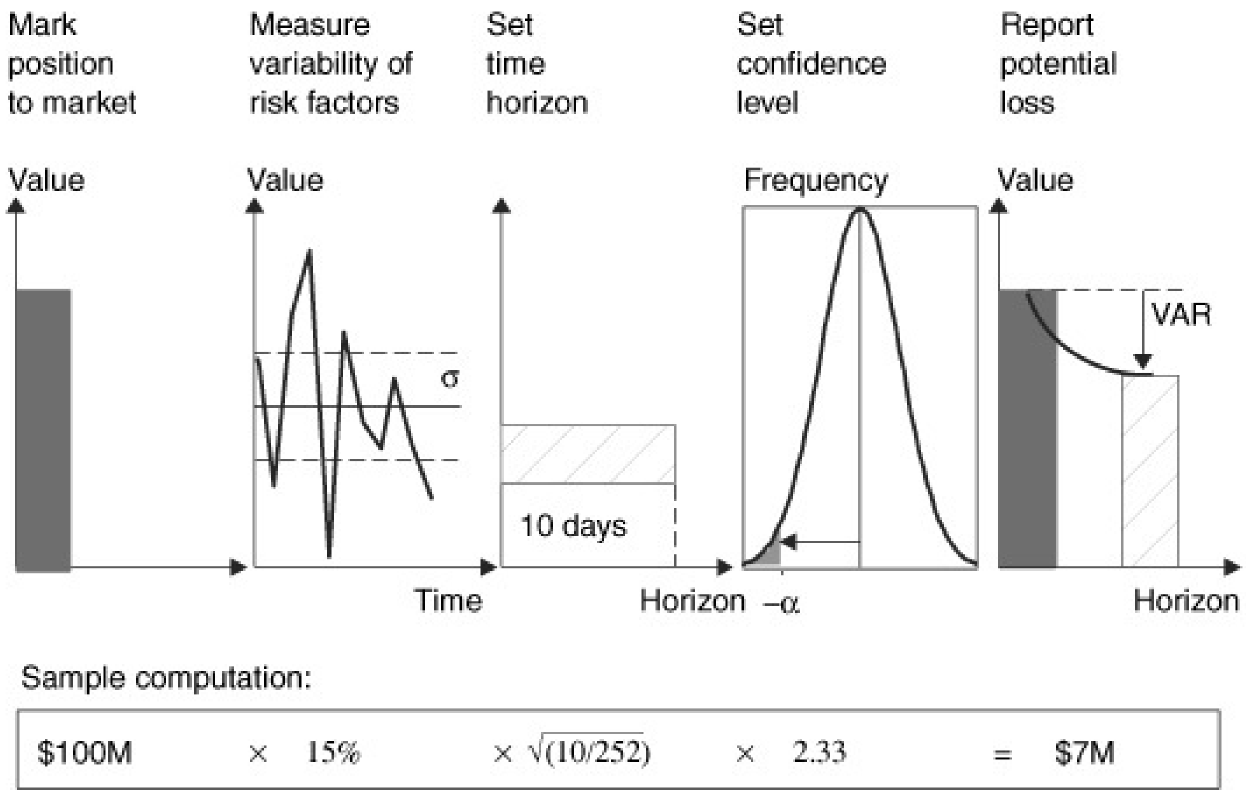
• *Set the confidence level* (e.g., 99 percent, which yields a 2.33 factor, assuming a normal distribution).

• *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.



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:

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

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

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

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.

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

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

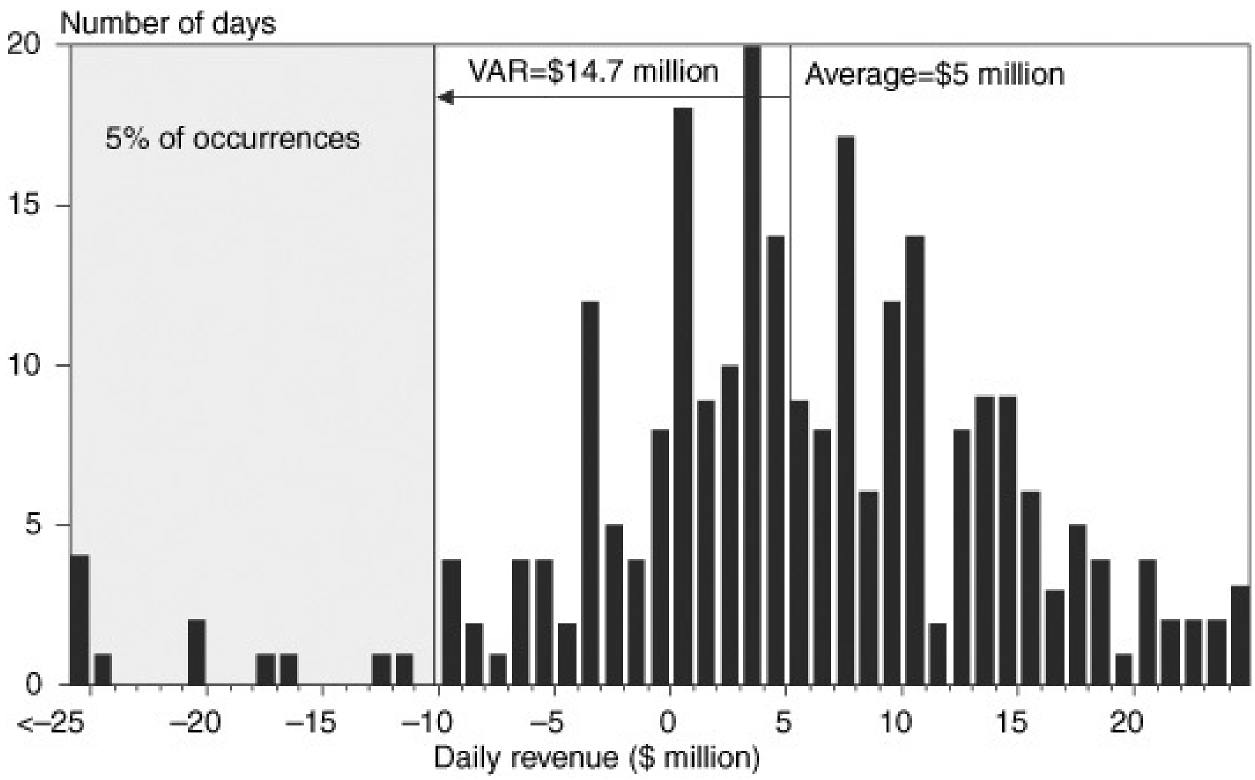
$$1 - c = \int_{-\infty}^{W^*} f(w)dw = P(w \le W^*) = p \tag{5.5}$$

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

FIGURE 5-2

Computation of 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 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 α = 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 α × (α W_0) = 1.645 × \$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.

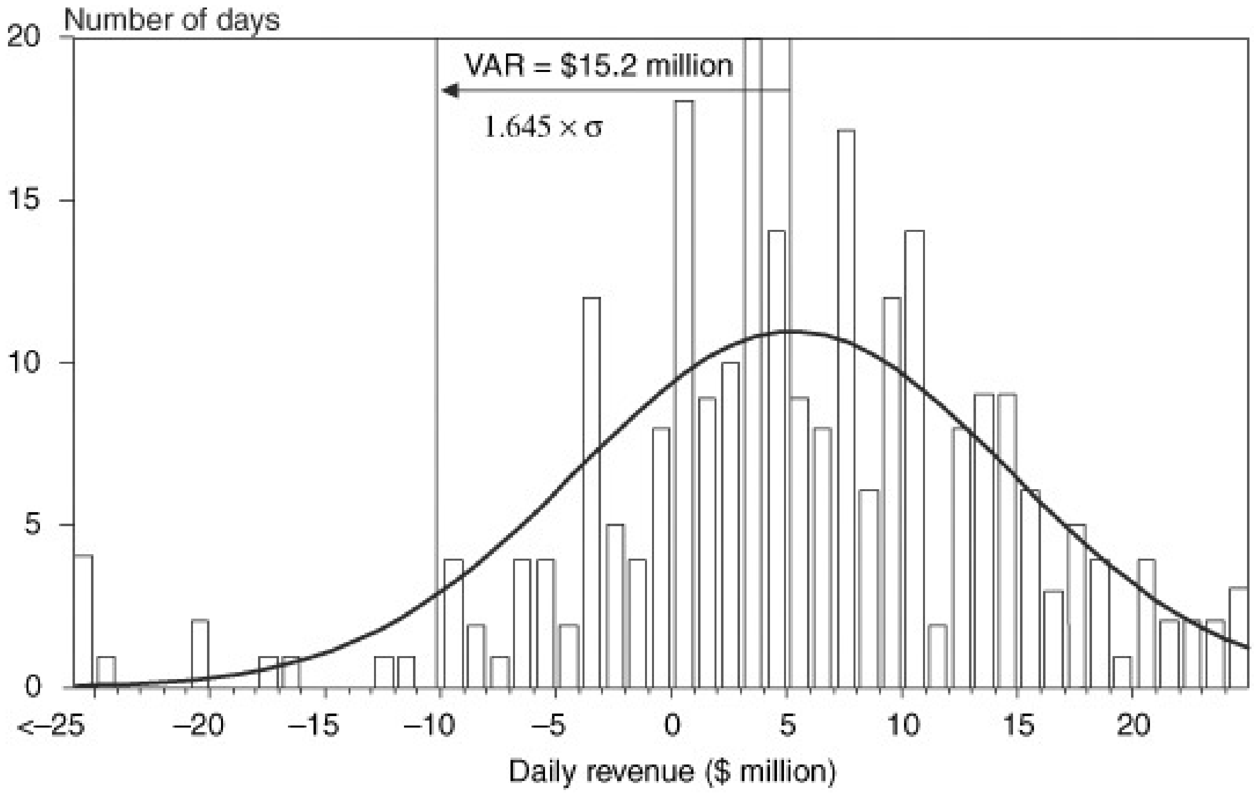
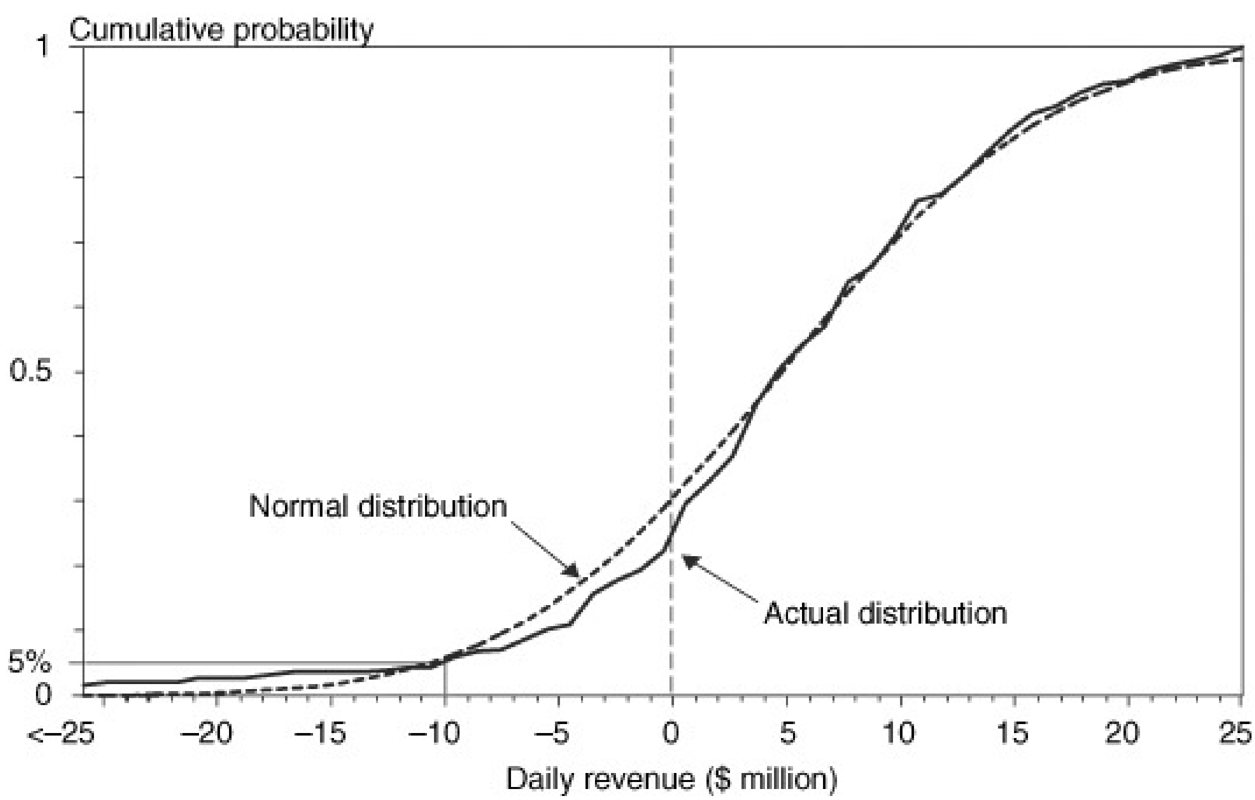


FIGURE 5-4
Comparison of cumulative distributions.



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 × $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 semi-deviation, 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:

- *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 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.

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.

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.

5.2.1 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 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.

TABLE 5-1
Credit Rating and Default Rates

Desired Rating	Default Rate 1 Year	Default Rate 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.

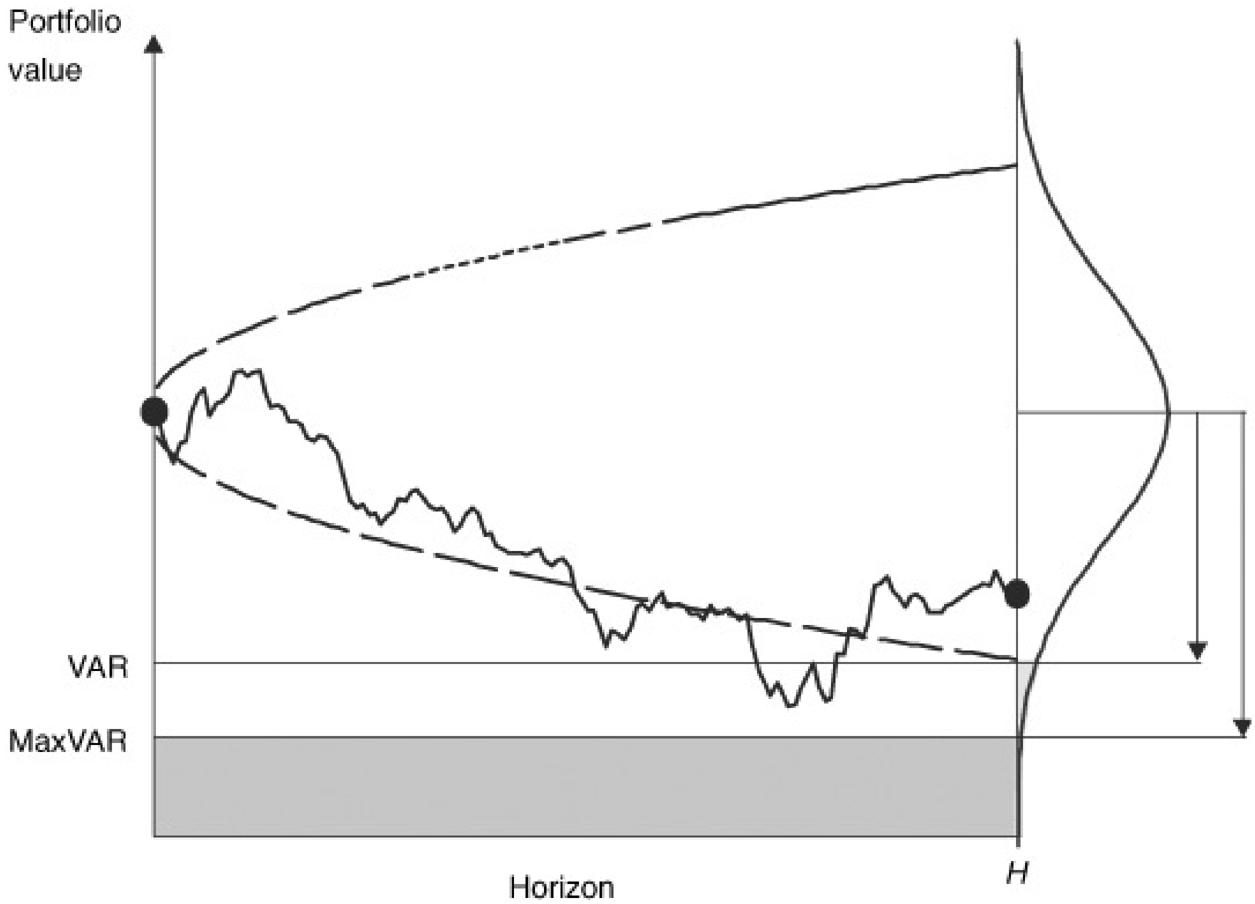


TABLE 5-2

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

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 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 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.

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.

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:

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

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.

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:

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

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 chisquare 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 chisquare 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

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

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.

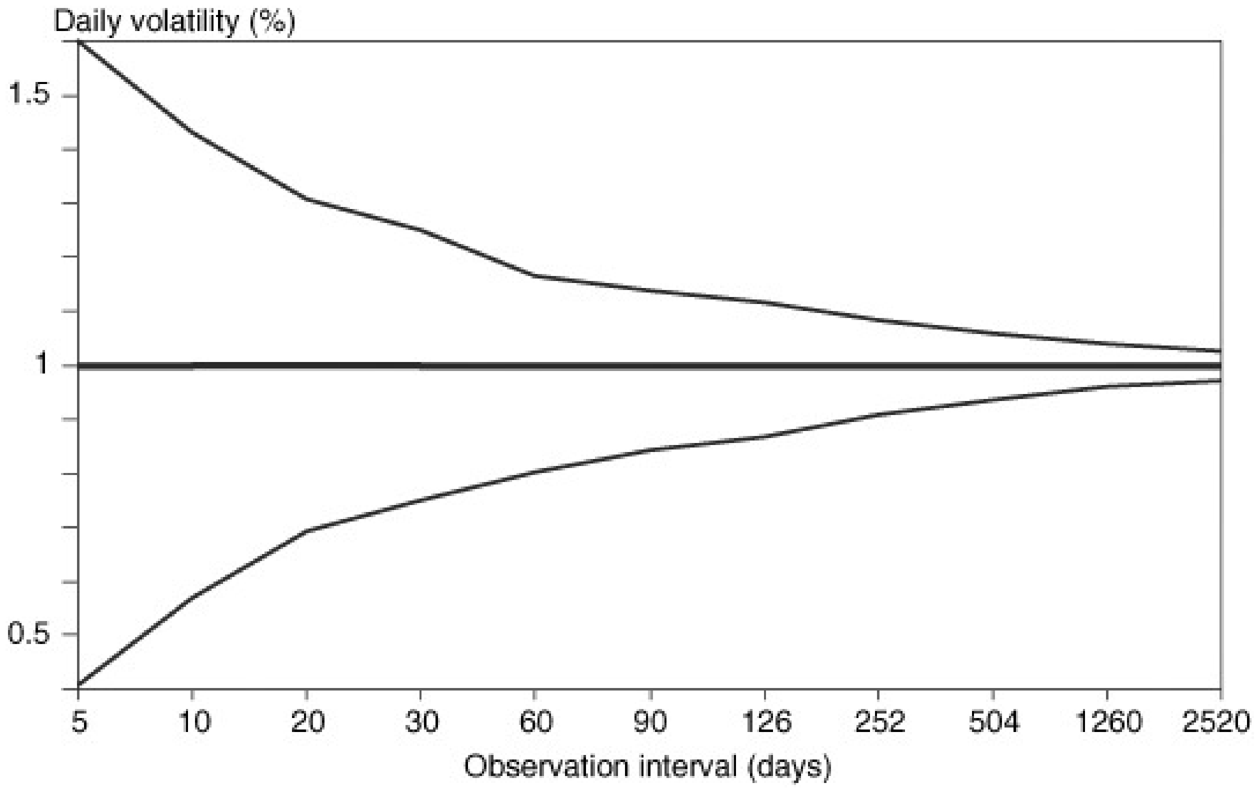
For instance, consider monthly returns on the euro/$ rate from 1973 to 2004. Sample parameters are $\hat{\mu}$ = −0.15 percent, $\hat{\sigma}$ = 3.39 percent, and T = 384 observations. The standard error of the estimate indicates how confident we are about the sample value; the smaller the error, the more confident we are. One standard error in $\hat{\mu}$ is $\mathrm{SE}(\hat{\mu}) = \hat{\sigma}\sqrt{1/T} = 3.39\sqrt{1/384} = 0.17$ 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 $\mathrm{SE}(\hat{\sigma}) = \hat{\sigma}\sqrt{1/2\mathrm{T}} = 3.39\sqrt{1/768} = 0.12$ 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 × 0.12, 3.39 + 1.96 × 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 *c*th 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

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

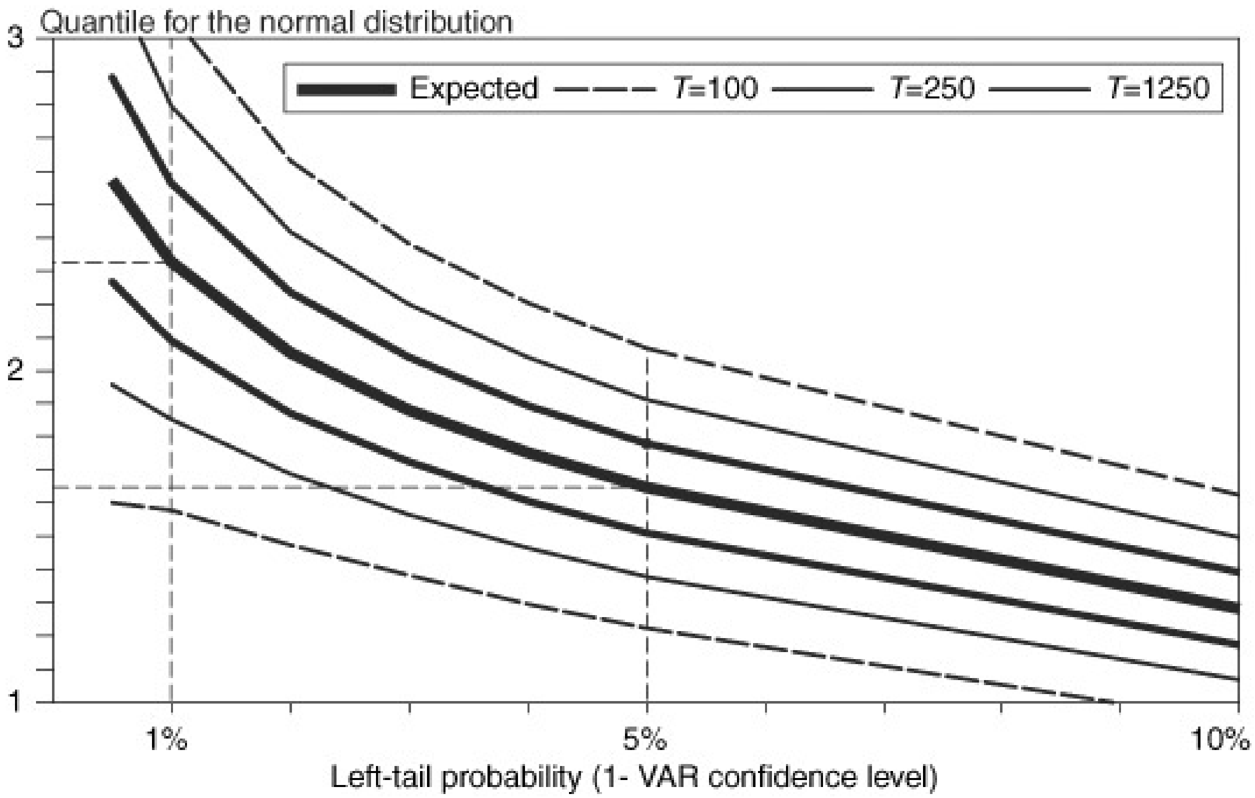
where T is the sample size, and $f(\cdot)$ is the probability distribution function evaluated at the quantile q. The effect of estimation error is illustrated in Figure 5-7, where the expected quantile and 95 percent confidence bands are plotted for quantiles from the normal distribution.

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 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.

FIGURE 5-7

Confidence bands for sample quantiles.



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

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]

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 $\mathrm{SE}(\hat{q}_\sigma) = \alpha \times \mathrm{SE}(\hat{\sigma}) = 1.65 \times (1 / \sqrt{2\times254}) \times \$9.2$ 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:

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

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 *c*th 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

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

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.

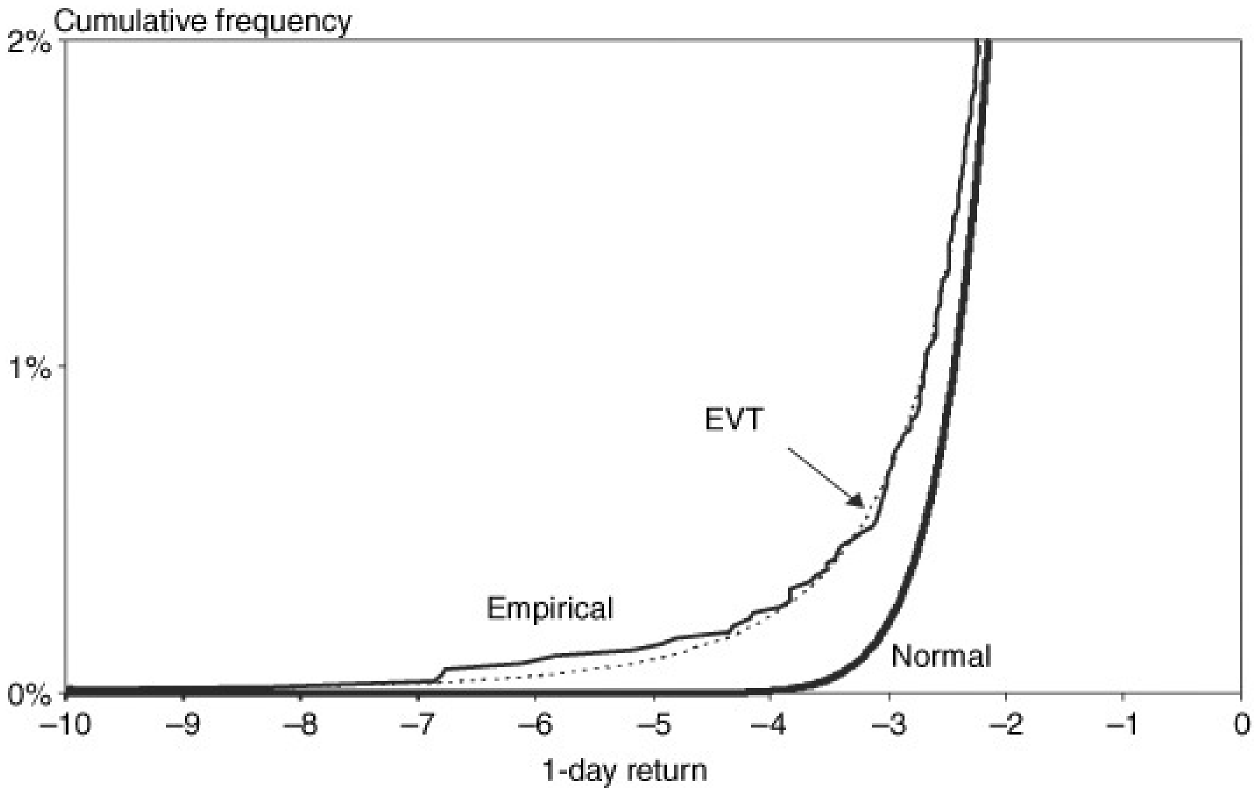


TABLE 5-5

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

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

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

This defines a maximum value $VAR_{max} = r\sigma$. 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

$$VAR_N = \alpha(99\%)\sigma = 2.326\sigma \tag{5.25}$$

If the true distribution is misspecified, the correction factor then is which happens to justify the correction factor applied by the Basel Committee.

$$k = \frac{VAR_{max}}{VAR_N} = \frac{7.071\sigma}{2.326\sigma} = 3.03 \tag{5.26}$$

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

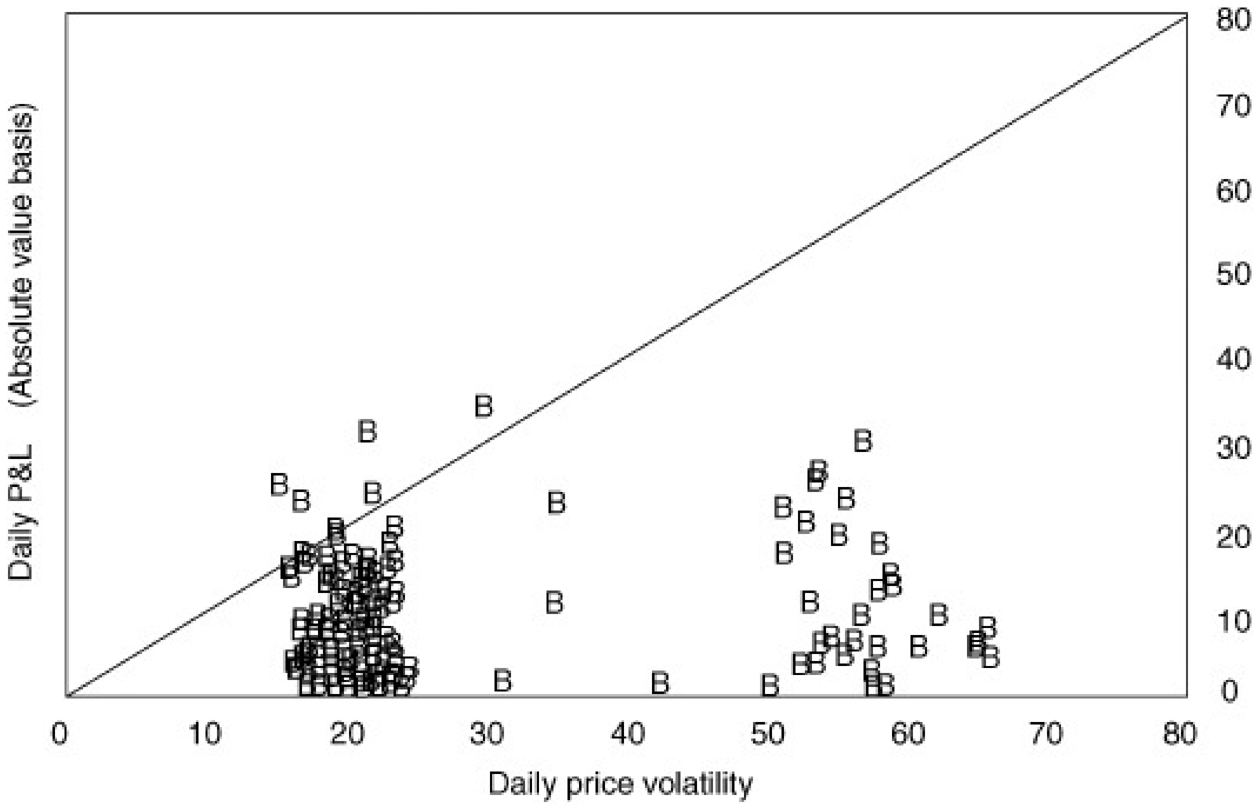
VAR models are only useful insofar as they can be demonstrated to be reasonably accurate. To do this, users must check systematically the validity of the underlying valuation and risk models through comparison of predicted and actual loss levels.

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 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.

FIGURE 6-1
Model evaluation: Bankers Trust.



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 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:

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

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

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

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.

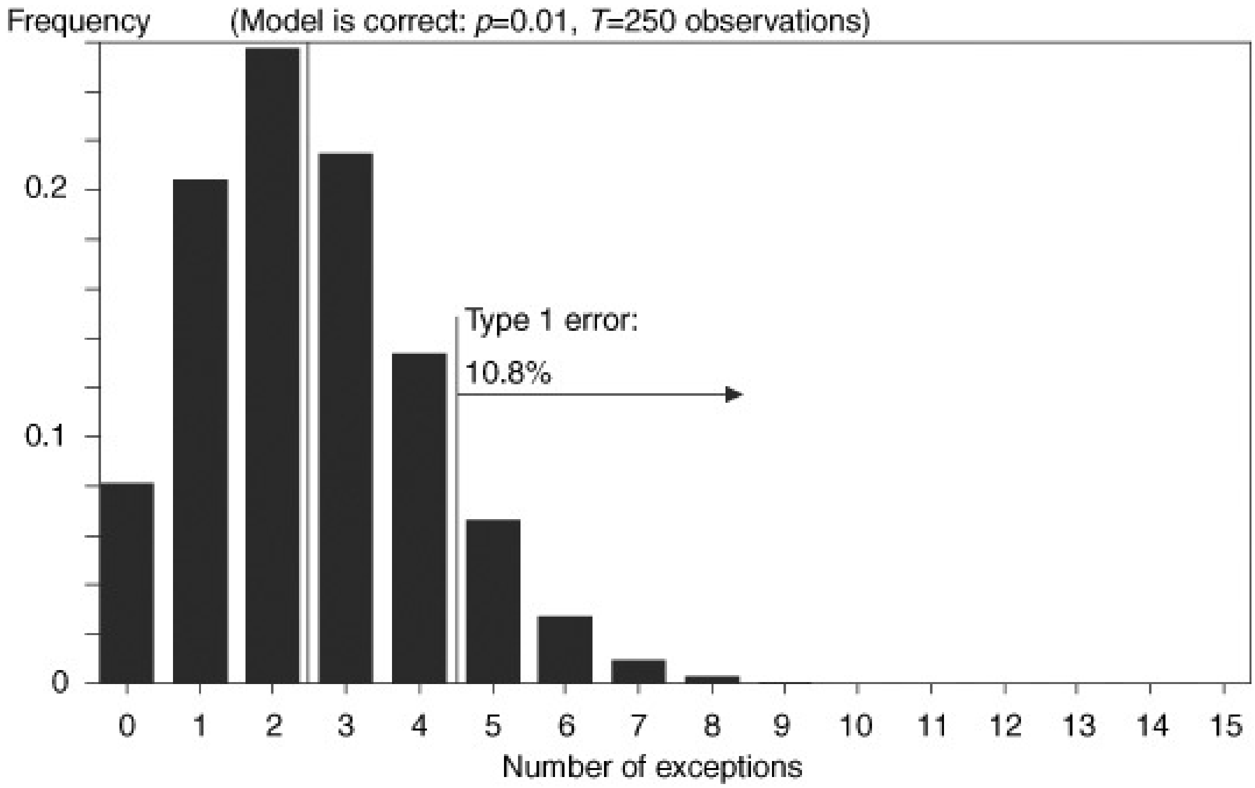
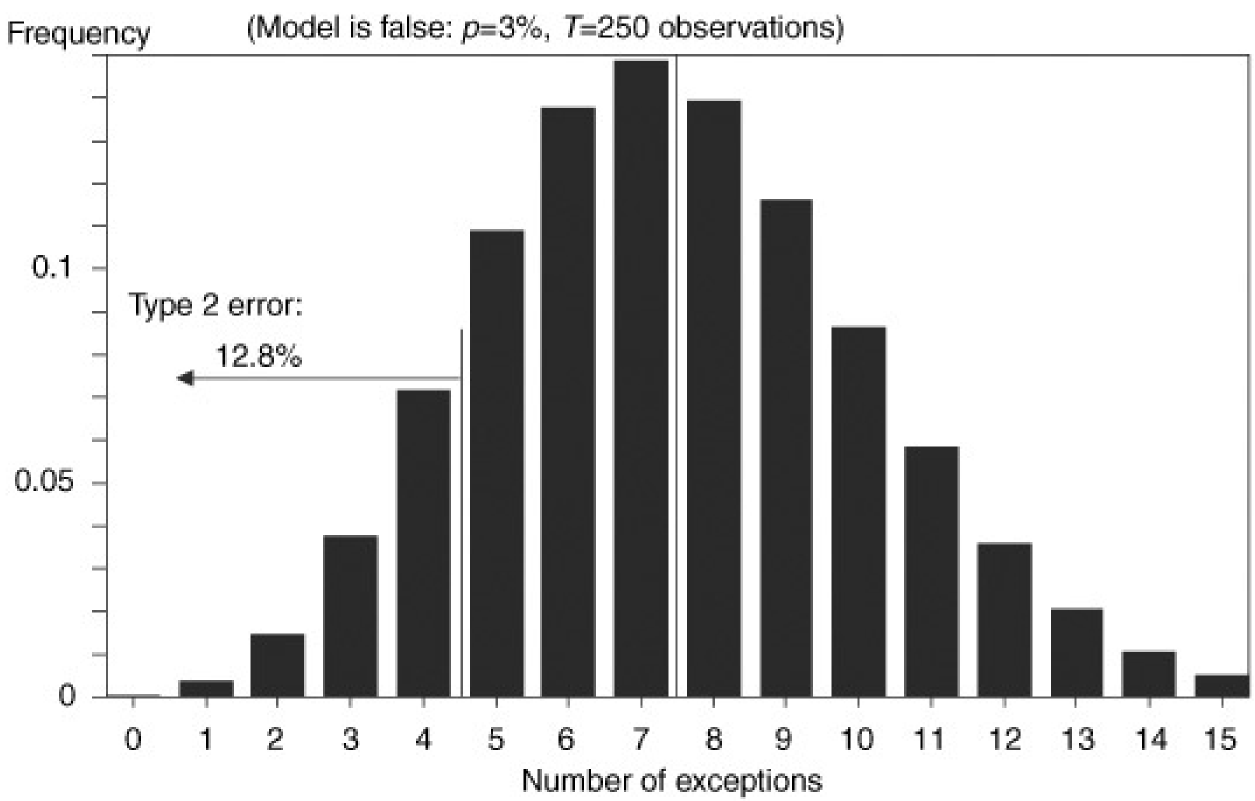


FIGURE 6-3

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: 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 $LR > 3.841$. This test is equivalent to Equation (6.2) because a chi-square variable is the square of a normal variable.

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

Probability level p	VAR Confidence Level c	Nonrejection Region for Number of Failures N		
		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).

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

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 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.

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

$$LR_{ind} = -2 \ln [(1-\pi)^{(T_{00}+T_{10})} \pi^{(T_{01}+T_{11})}] + 2 \ln [(1-\pi_0)^{T_{00}} \pi_0^{T_{01}} (1-\pi_1)^{T_{10}} \pi_1^{T_{11}}] \tag{6.4}$$

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.

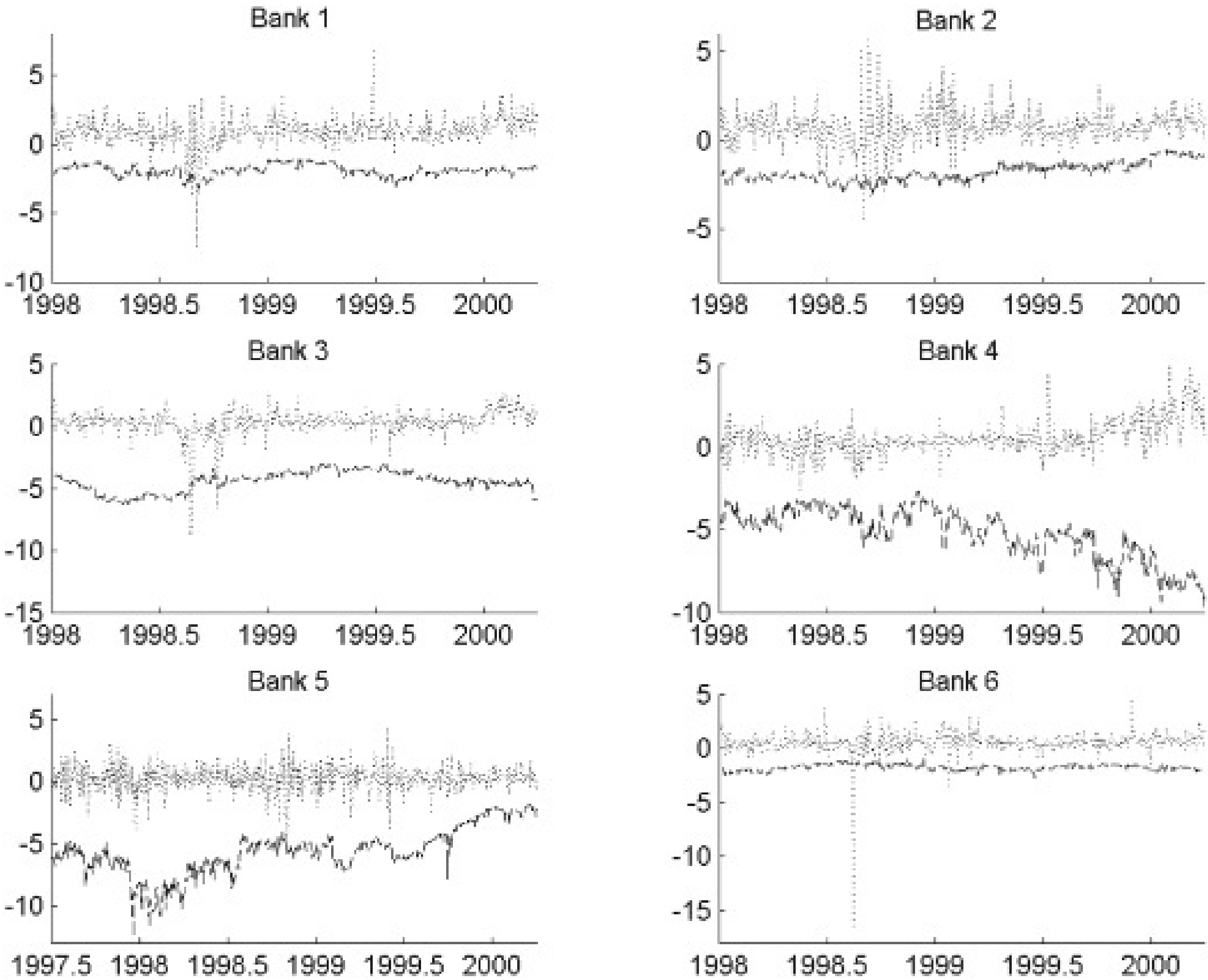


Figure 6-4 plots the time series of P&L along with the daily VAR (the lower lines) for a sample of six U.S. commercial banks. With approximately 600 observations, we should observe on average 6 violations, given a VAR confidence level of 99 percent.

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 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 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.

$$R_{p,t+1} = \sum_{i=1}^{N} w_i R_{i,t+1} \tag{7.1}$$

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:

$$R_p = w_1 R_1 + w_2 R_2 + \cdots + w_N R_N = [w_1 w_2 \cdots w_N] \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix} = w' R \qquad (7.2)$$

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 and the variance is

$$E(R_p) = \mu_p = \sum_{i=1}^{N} w_i \mu_i \qquad (7.3)$$

$$V(R_p) = \sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} w_i w_j \sigma_{ij} = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j<i}^{N} w_i w_j \sigma_{ij} \qquad (7.4)$$

This sum accounts not only for the risk of the individual securities σ^2_i 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

$$\sigma_p^2 = [w_1 \cdots w_N] \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \ldots & \sigma_{1N} \\ \vdots & & & & \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \ldots & \sigma_N^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}$$

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 − α 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.

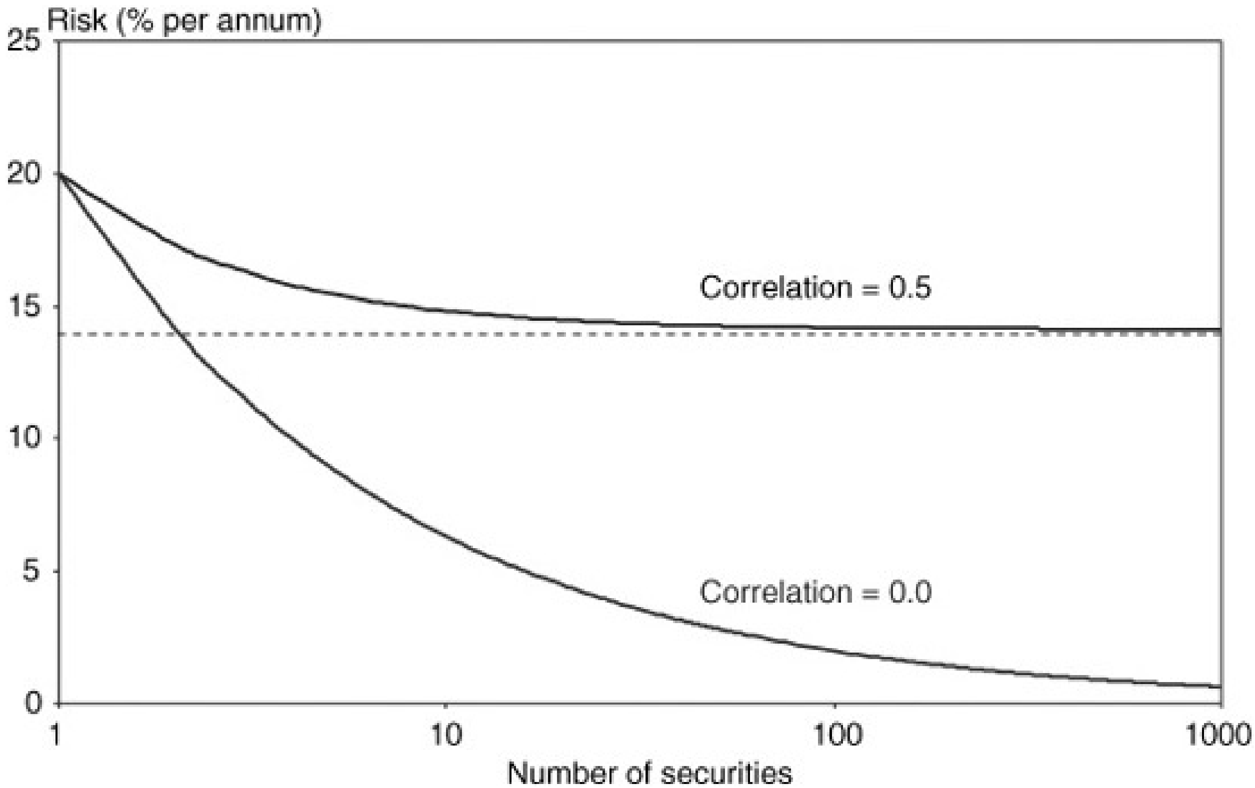
Equation (7.4) shows that the portfolio VAR depends on variances, covariances, and the number of assets. Covariance is a measure of the extent to which two variables move linearly together. If two variables are independent, their covariance is equal to zero. A positive covariance means that the two variables 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

$$\sigma_p = \sigma\sqrt{\frac{1}{N} + \left(1 - \frac{1}{N}\right)\rho} \qquad (7.10)$$

which tends to $\sigma\sqrt{\rho}$ as *N* increases. Thus, when ρ = 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

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2 \qquad (7.11)$$

The portfolio VAR is then

$$\mathrm{VAR}_p = \alpha\sigma_p W = \alpha\sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}\,W \qquad (7.12)$$

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

$$\mathrm{VAR}_p = \sqrt{\alpha^2 w_1^2 W^2 \sigma_1^2 + \alpha^2 w_2^2 W^2 \sigma_2^2} = \sqrt{\mathrm{VAR}_1^2 + \mathrm{VAR}_2^2} \qquad (7.13)$$

The portfolio risk must be lower than the sum of the individual VARs: $\mathrm{VAR}_p < \mathrm{VAR}_1 + \mathrm{VAR}_2$. 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

$$\mathrm{VAR}_p = \sqrt{\mathrm{VAR}_1^2 + \mathrm{VAR}_2^2 + 2\mathrm{VAR}_1 \times \mathrm{VAR}_2} = \mathrm{VAR}_1 + \mathrm{VAR}_2 \qquad (7.14)$$

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 α 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,

$$\frac{\partial \sigma_p^2}{\partial w_i} = 2w_i\sigma_i^2 + 2\sum_{j=1, j\neq i}^{N} w_j\sigma_{ij} = 2\operatorname{cov}(R_i, w_iR_i + \sum_{j\neq i}^{N} w_jR_j) = 2\operatorname{cov}(R_i, R_p) \quad (7.15)$$

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

$$\frac{\partial \sigma_p}{\partial w_i} = \frac{\operatorname{cov}(R_i, R_p)}{\sigma_p} \quad (7.16)$$

Converting into a VAR number, we find an expression for the *marginal VAR*, which is a vector with component

$$\Delta \text{VAR}_i = \frac{\partial \text{VAR}}{\partial x_i} = \frac{\partial \text{VAR}}{\partial w_i W} = \alpha\frac{\partial \sigma_p}{\partial w_i} = \alpha\frac{\operatorname{cov}(R_i, R_p)}{\sigma_p} \quad (7.17)$$

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

$$\beta_i = \frac{\mathrm{cov}(R_i, R_p)}{\sigma_p^2} = \frac{\sigma_{ip}}{\sigma_p^2} = \frac{\rho_{ip}\sigma_i\sigma_p}{\sigma_p^2} = \rho_{ip}\frac{\sigma_i}{\sigma_p} \qquad (7.18)$$

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,

$$R_{i,t} = \alpha_i + \beta_i R_{p,t} + \epsilon_{i,t} \qquad t = 1, \ldots, T \qquad (7.19)$$

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

$$\Delta\mathrm{VAR}_i = \frac{\partial\mathrm{VAR}}{\partial x_i} = \alpha(\beta_i \times \sigma_p) = \frac{\mathrm{VAR}}{W} \times \beta_i \qquad (7.20)$$

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

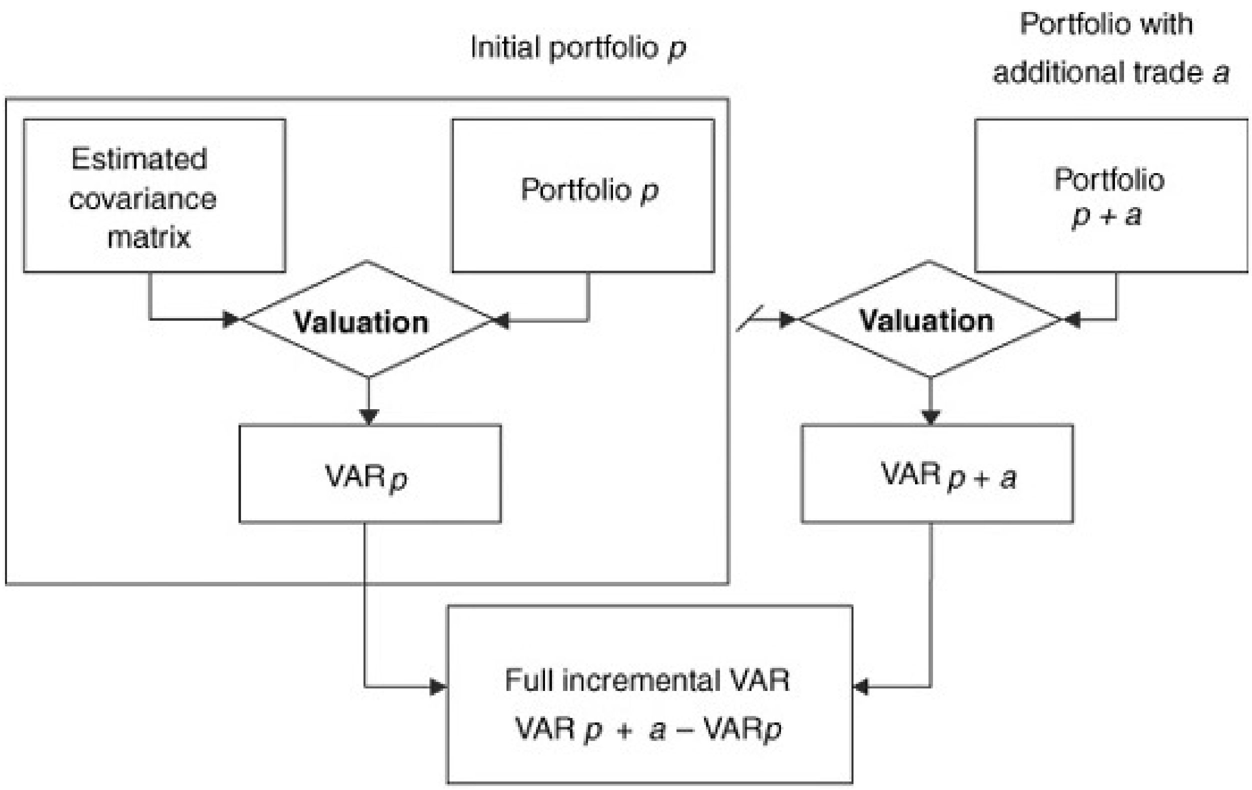
$$\text{Incremental VAR} = \text{VAR}_{p+a} - \text{VAR}_p \tag{7.21}$$

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,

$$\mathrm{VAR}_{p+a} = \mathrm{VAR}_p + (\Delta\mathrm{VAR})' \times a + \cdots \tag{7.22}$$

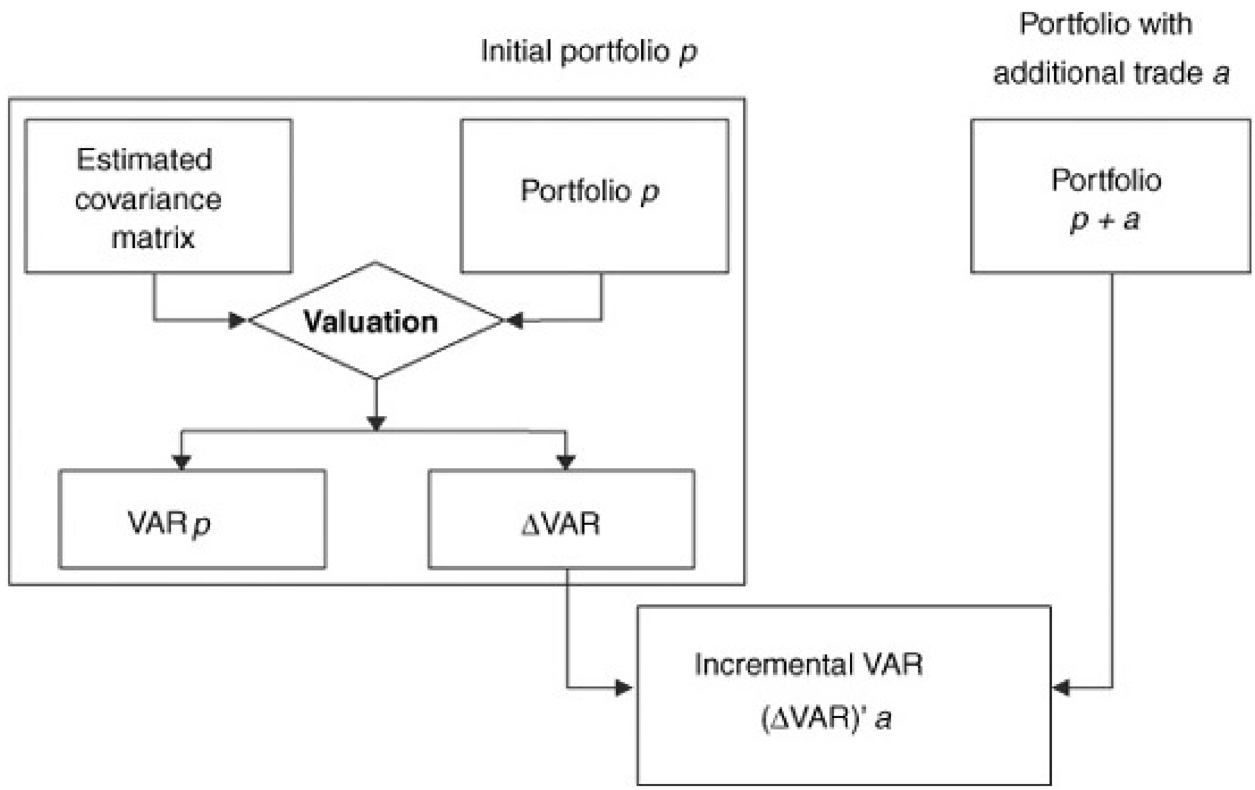
where we ignored second-order terms if the deviations a are small. Hence the incremental VAR can be reported as, approximately,

$$\text{Incremental VAR} \approx (\Delta\mathrm{VAR})' \times a \tag{7.23}$$

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.

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

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

$$\sigma^2_{p+a}\text{W}^2_{p+a} = [\$2.01\ \$1]\begin{bmatrix}0.05^2 & 0\\0 & 0.12^2\end{bmatrix}\begin{bmatrix}\$2.01\\\$1\end{bmatrix}$$

which gives VAR_{p+a} = $258,267. Relative to the initial 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 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 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.

$$\text{CVAR}_1 + \text{CVAR}_2 + \cdots + \text{CVAR}_N = \text{VAR}\left(\sum_{i=1}^{N} w_i \beta_i\right) = \text{VAR} \quad (7.28)$$

> **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

$$\text{CVAR}_i = \text{VAR} w_i \beta_i = (\alpha \sigma_p W) w_i \beta_i = (\alpha \sigma_i w_i W) \rho_i = \text{VAR}_i \rho_i \quad (7.29)$$

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

$$\text{Percent contribution to VAR of component } i = \frac{\text{CVAR}_i}{\text{VAR}} = w_i \beta_i \quad (7.30)$$

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

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

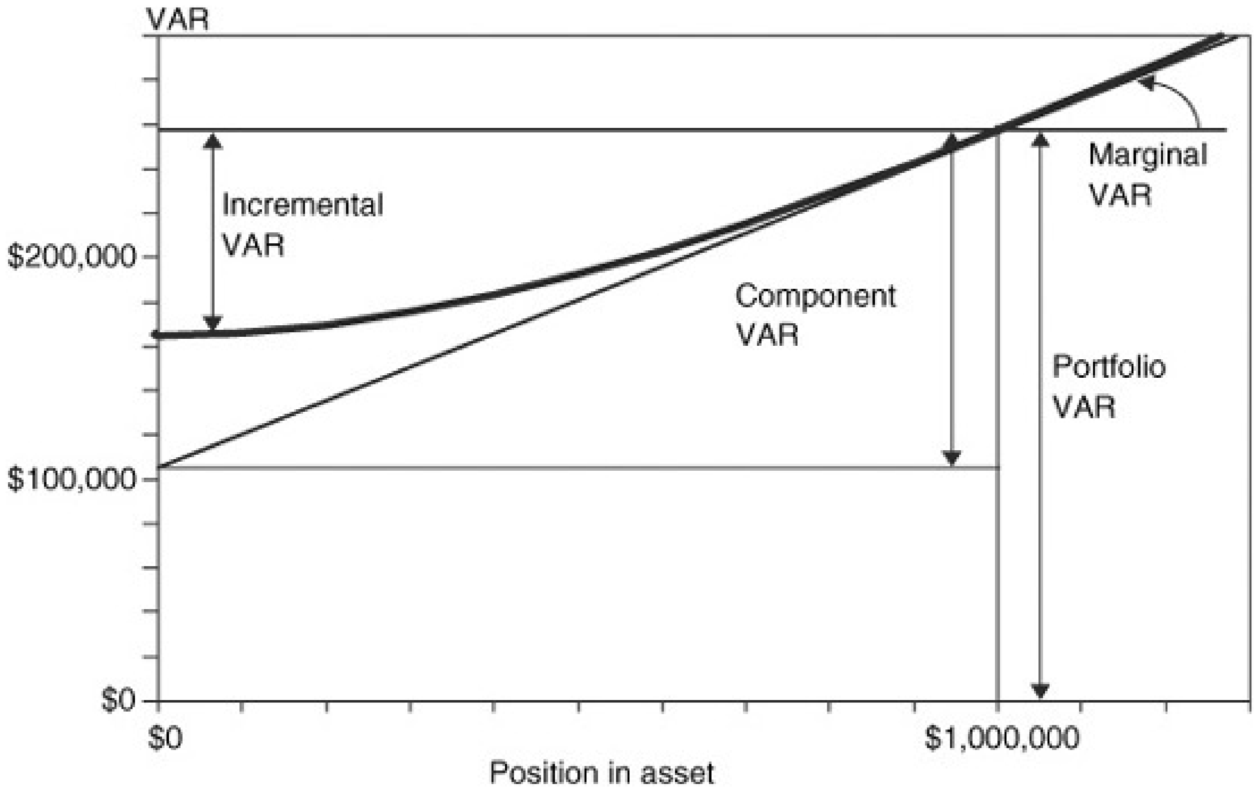
$$\begin{bmatrix} \text{CVAR}_1 / \text{VAR} \\ \text{CVAR}_2 / \text{VAR} \end{bmatrix} = \begin{bmatrix} w_1\beta_1 \\ w_2\beta_2 \end{bmatrix} = \begin{bmatrix} 0.667 \times 0.615 \\ 0.333 \times 1.770 \end{bmatrix} = \begin{bmatrix} 41.0\% \\ 59.0\% \end{bmatrix}$$

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_iW	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 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).

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 $\sigma^2_1 x_1 + \sigma_{12} x_2 = 0.000139 \times (-\$16{,}000) + (-0.000078) \times \$7700 = -2.82$. 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 × \$506, or \$835 million.

TABLE 7-3

Barings’ Risks

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.

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

$$R_p^* = \sum_{i=1}^{N} E(R_i | R_p = R_p^*) w_i \quad (7.34)$$

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.

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:

$$\Delta\text{VAR}_i = \frac{\text{VAR}}{W} \times \beta_i = \text{constant} \quad (7.35)$$

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.

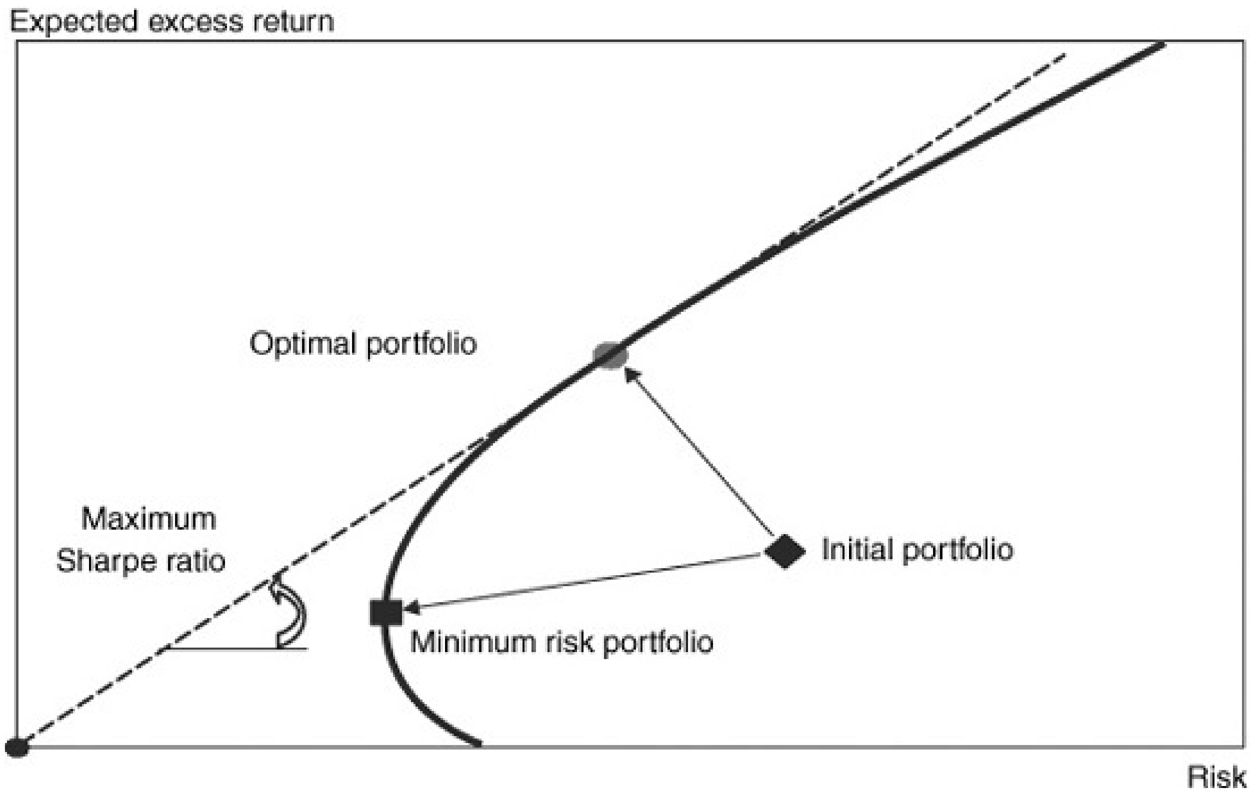
7.5.2 From Risk Management to Portfolio Management

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

$$\mathrm{SR}_p = \frac{E_p}{\sigma_p} \tag{7.37}$$

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,

$$\frac{E_i}{\Delta \mathrm{VAR}_i} = \frac{E_i}{\beta_i} = \text{constant} \tag{7.38}$$

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 *i*th row of *A* individually by each element of the *j*th 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.

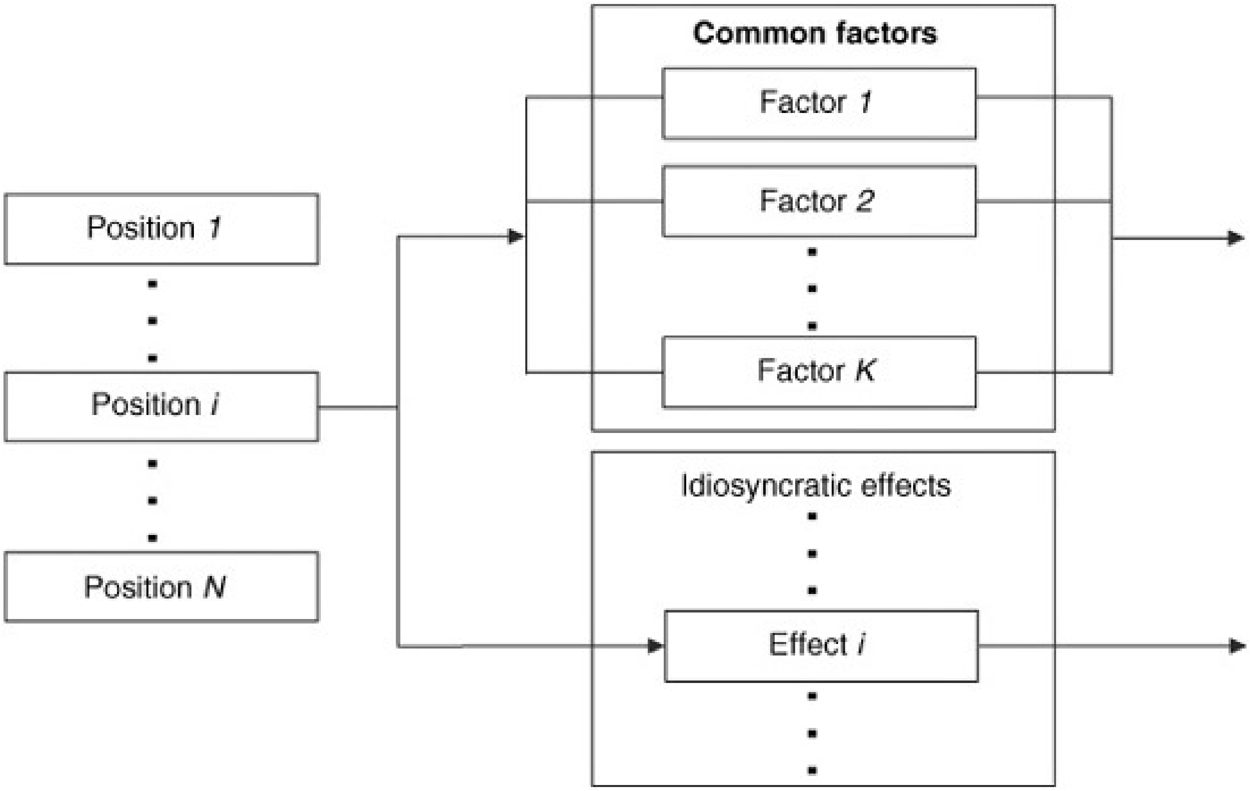
Problems also occur when the series are linearly correlated. This happens, for example, when two assets are identical ($\rho = 1$). In this situation, a portfolio consisting of \$1 on the first asset and −\$1 on the second will have exactly zero risk. In practice, this problem is more likely to occur 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 a full covariance matrix. In addition, it lends itself better to an economic intuition, which helps to understand the results.

FIGURE 8-1
Simplifying the risk structure.



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 $\text{cov}(\epsilon_i, R_m) = E(\epsilon_i R_m) - E(\epsilon_i) E(R_m) = 0$. 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

$$\begin{aligned} \sigma_i^2 = V(\beta_i R_m + \epsilon_i) &= \beta_i^2 \sigma_m^2 + 2\,\text{cov}(\beta_i R_m, \epsilon_i) + V(\epsilon_i) \\ &= \beta_i^2 \sigma_m^2 + \sigma^2{}_{\epsilon,i} \end{aligned} \tag{8.4}$$

because R_m and ϵ_i are uncorrelated. The covariance between two assets is which is solely due to the common factor because all the other terms are zero owing to Equation (8.3).

$$\sigma_{i,j} = \text{cov}(\beta_i R_m + \epsilon_i, \beta_j R_m + \epsilon_j) = \beta_i \beta_j \sigma_m^2 \tag{8.5}$$

As a result, we can construct the full covariance matrix as

$$\Sigma = \begin{bmatrix} \beta_1\beta_1\sigma_m^2 + \sigma_{\varepsilon,1}^2 & \cdots & \beta_1\beta_N\sigma_m^2 \\ \vdots & & \vdots \\ \beta_N\beta_1\sigma_m^2 & \cdots & \beta_N\beta_N\sigma_m^2 + \sigma_{\varepsilon,N}^2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} \left[\beta_1 \cdots \beta_N\right]\sigma_m^2 + \begin{bmatrix} \sigma_{\varepsilon,1}^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_{\varepsilon,N}^2 \end{bmatrix}$$

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

$$V(R_p) \to (w'\beta\beta'w)\ \sigma_m^2 = (\beta_p\sigma_m)^2 \tag{8.9}$$

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 $\beta_1\beta_2V(R_m)$, 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,

$$R_i = \alpha_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i \qquad (8.10)$$

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,

$$\Sigma = \beta_1 \beta_1' \sigma_1^2 + \cdots + \beta_K \beta_K' \sigma_K^2 + D_\epsilon \qquad (8.11)$$

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

$$V(R_p) \rightarrow (\beta_{1p}\sigma_1)^2 + \ldots + (\beta_{Kp}\sigma_K)^2 \qquad (8.12)$$

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

$$\sigma(dP/P) = |D^*| \times \sigma(dy) \tag{8.14}$$

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

$$\mathrm{VAR}(dP/P) = |D^*| \times \mathrm{VAR}(dy) \qquad (8.15)$$

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

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.

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 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.

$$x = D^* \times \sigma(dy) \times P = D^* \times [\mathrm{VAR}(dy)/1.65] \times \mathrm{P} \qquad (8.17)$$

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 largely hedged against the first factor, with $\beta_{1p} = -0.019$ only. The two-factor risk analysis gives

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				$\beta_{1p} = +2.017$		$\beta_{1p} = -0.019$
				$\beta_{2p} = -3.005$		$\beta_{2p} = +0.829$

$$\sigma^2(R_p) = (-0.019^2 \times 10.104) + (0.829^2 \times 0.662) = 0.004 + 0.455 = 0.459$$

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]

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.

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,

$$\text{VAR}_1 = \alpha W \sigma_m$$

- *Beta mapping* only considers the net beta of the portfolio, that is,

$$\text{VAR}_2 = \alpha W (\beta_p \sigma_m)$$

- *Diagonal model* considers both the beta and specific risk, that is,

$$\text{VAR}_3 = \alpha W \sqrt{(\beta_p \sigma_m)^2 + w' D_\epsilon w}$$

- *Industry mapping* replaces each stock by a like position in an industry index I, that is,

$$\text{VAR}_4 = \alpha W \sqrt{w_I' \Sigma_I w_I}$$

- *Individual mapping* uses the full covariance matrix of individual stocks and provides an exact VAR measure over this sample period, that is,

$$\text{VAR}_5 = \alpha W \sqrt{w' \Sigma w}$$

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.

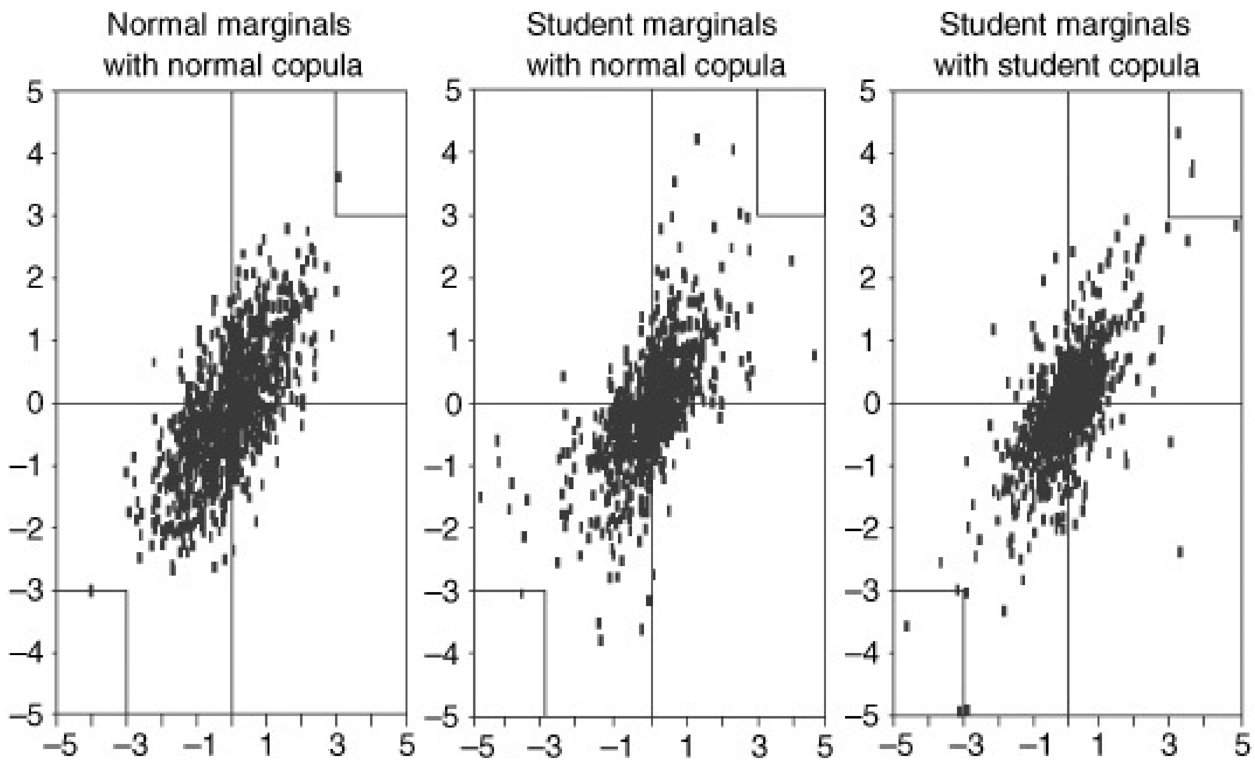


Figure 8-2 describes a plot of two variables generated with (1) normal marginals with a normal copula, (2) student marginals with a normal copula, and (3) student marginals with a student copula parametrized with 3 degrees of freedom. Now observe the boxes including extreme observations, either both 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 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*.

$$z_1 = \beta_{11}R_1 + \cdots + \beta_{N1}R_N = \beta_1'R \tag{8.23}$$

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

$$\Sigma^* = [\beta_1 \cdots \beta_K] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_K \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_K' \end{bmatrix} = \beta_1 \beta_1' \lambda_1 + \cdots + \beta_K \beta_K' \lambda_K \qquad (8.27)$$

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.

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): 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.

$$\begin{aligned} \sigma^2(R_p) &= w' \Sigma^* w = w' \beta_1 \beta_1' w \lambda_1 + \cdots + w' \beta_K \beta_K' w \lambda_K \\ &= (w' \beta_1)^2 \lambda_1 + \cdots + (w' \beta_K)^2 \lambda_K \qquad (8.28) \\ &= \delta_1^2 \sigma^2(z_1) + \cdots + \delta_K^2 \sigma^2(z_K) \end{aligned}$$

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 ∈ is 1; the residuals ∈ 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.

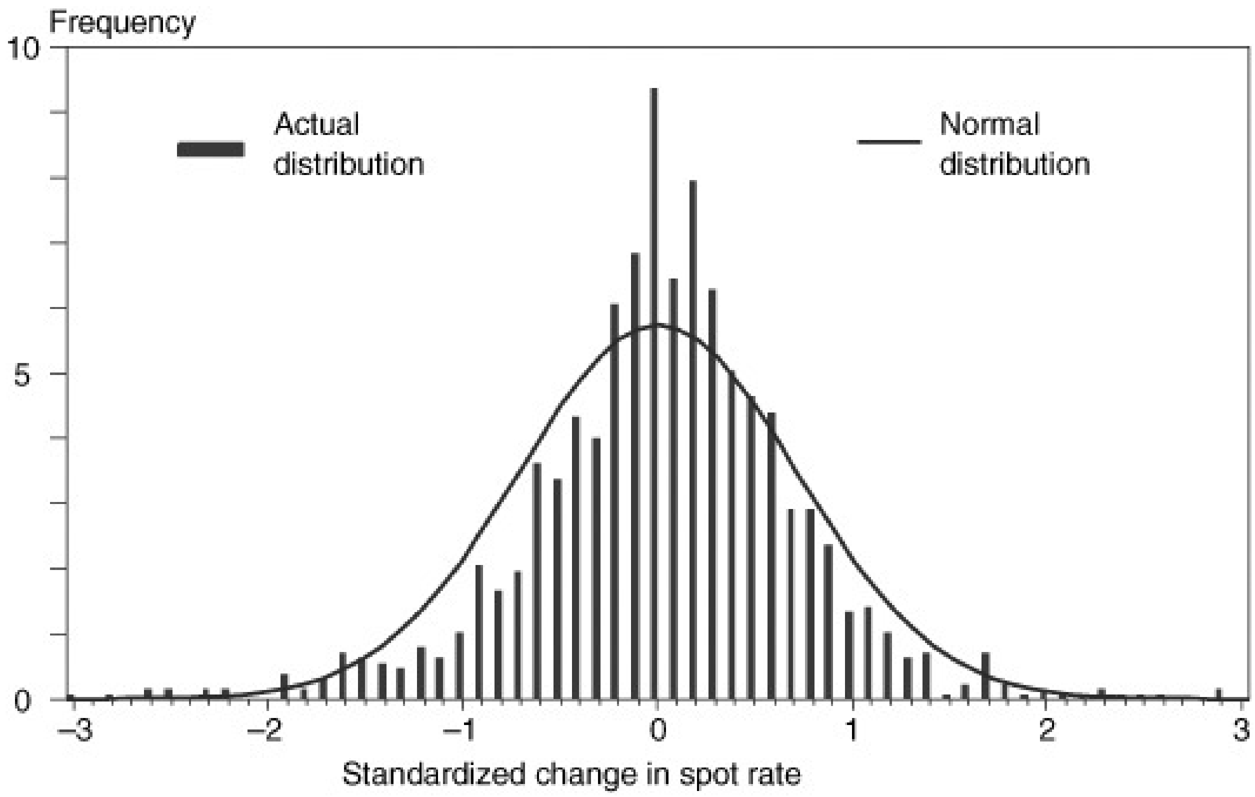
Figure 9-2 compares the normal approximation with the actual empirical distribution of the $/BP exchange rate. Relative to the normal model, the actual distribution contains more observations in the center and in the tails.

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 this chapter we focus on traditional approaches based on *parametric* time-series modeling.[1]

FIGURE 9-2

Distribution of the $/BP rate.



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.

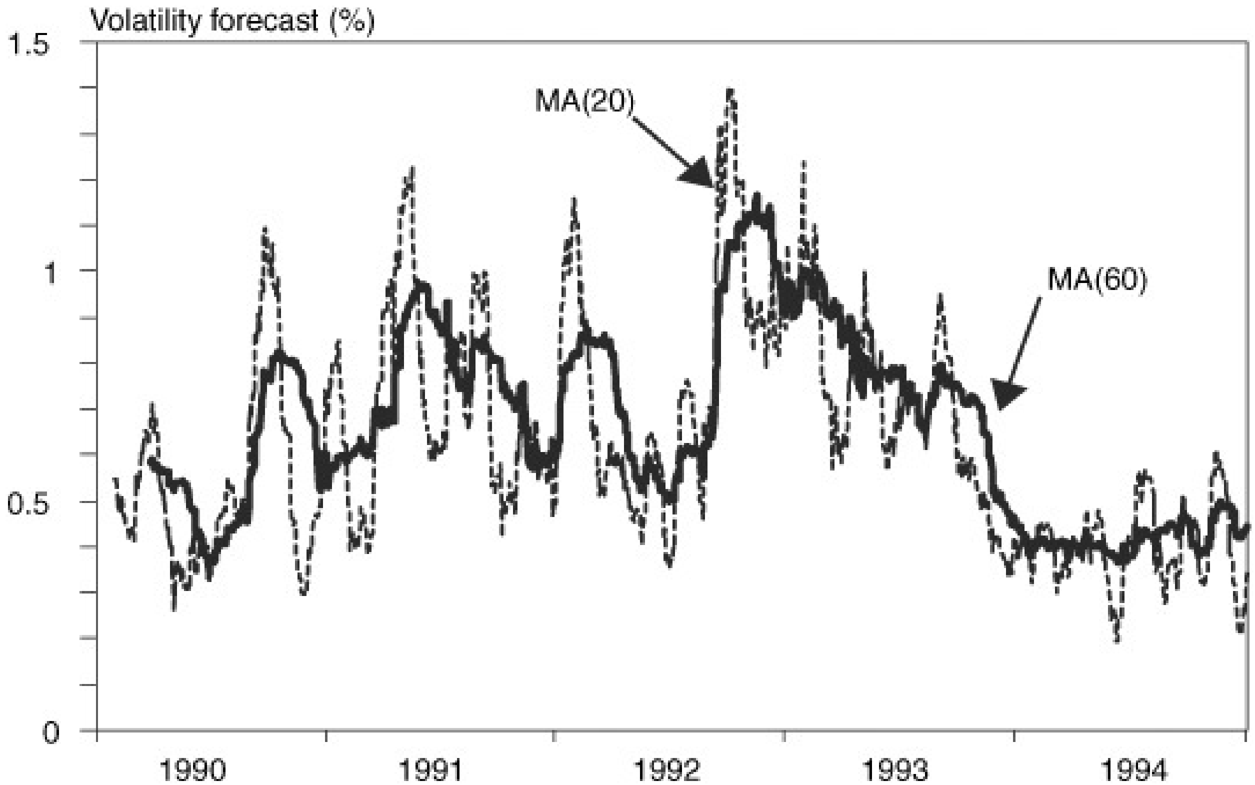
Each day, the forecast is updated by adding information from the preceding day and dropping information from (M + 1) days ago. All weights on past returns are equal and set to (1/M). Figure 9-3 displays 20-and 60-day moving averages for our $/BP rate.

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.

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

$$h = \frac{\alpha_0}{1 - \alpha_1 - \beta} \qquad (9.3)$$

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 for each time period t. The optimization maximizes the logarithm of the likelihood function, that is,

$$\max F(\alpha_0, \alpha_1, \beta \mid r) = \sum_{t=1}^{T} \ln f(r_t \mid h_t) = \sum_{t=1}^{T} \left(\ln \frac{1}{\sqrt{2\pi h_t}} - \frac{r_t^2}{2h_t} \right) \quad (9.4)$$

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,

$$\epsilon_t = \frac{r_t}{\sqrt{h_t}} \quad (9.5)$$

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

FIGURE 9-4

GARCH volatility forecast.

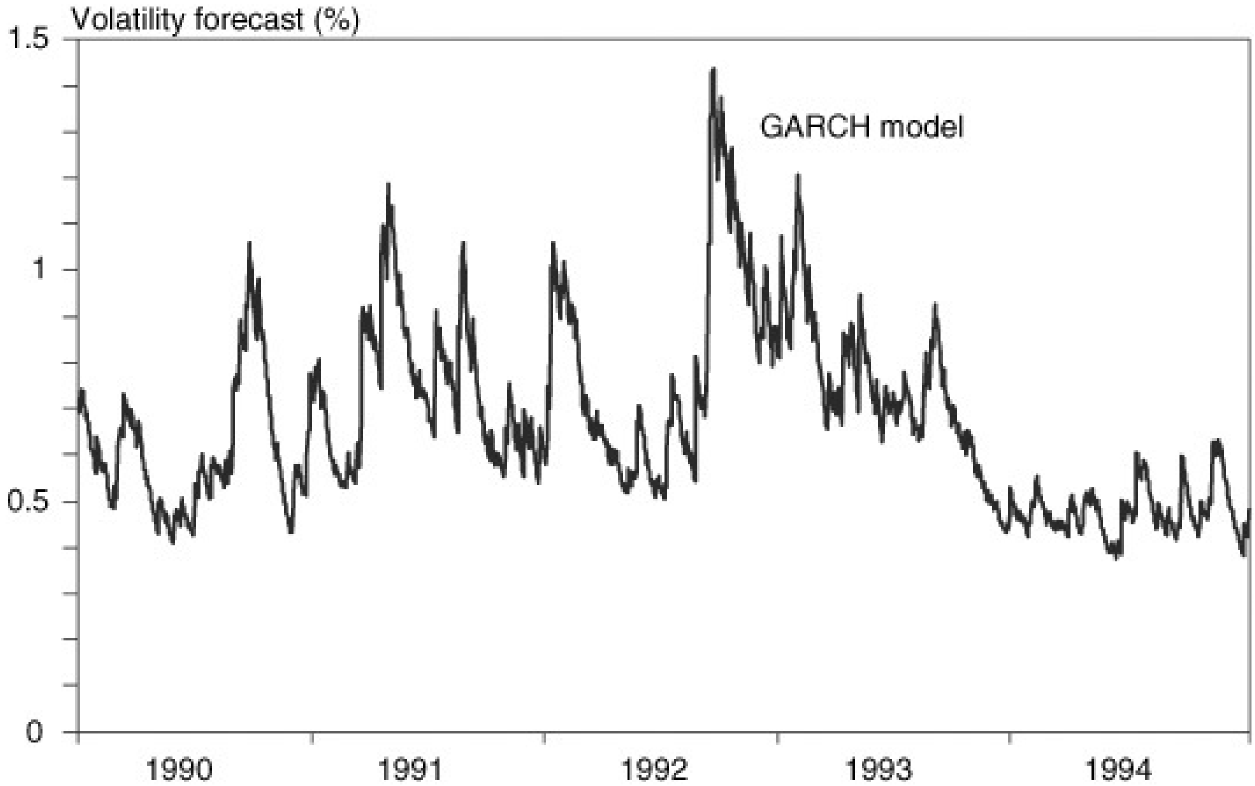


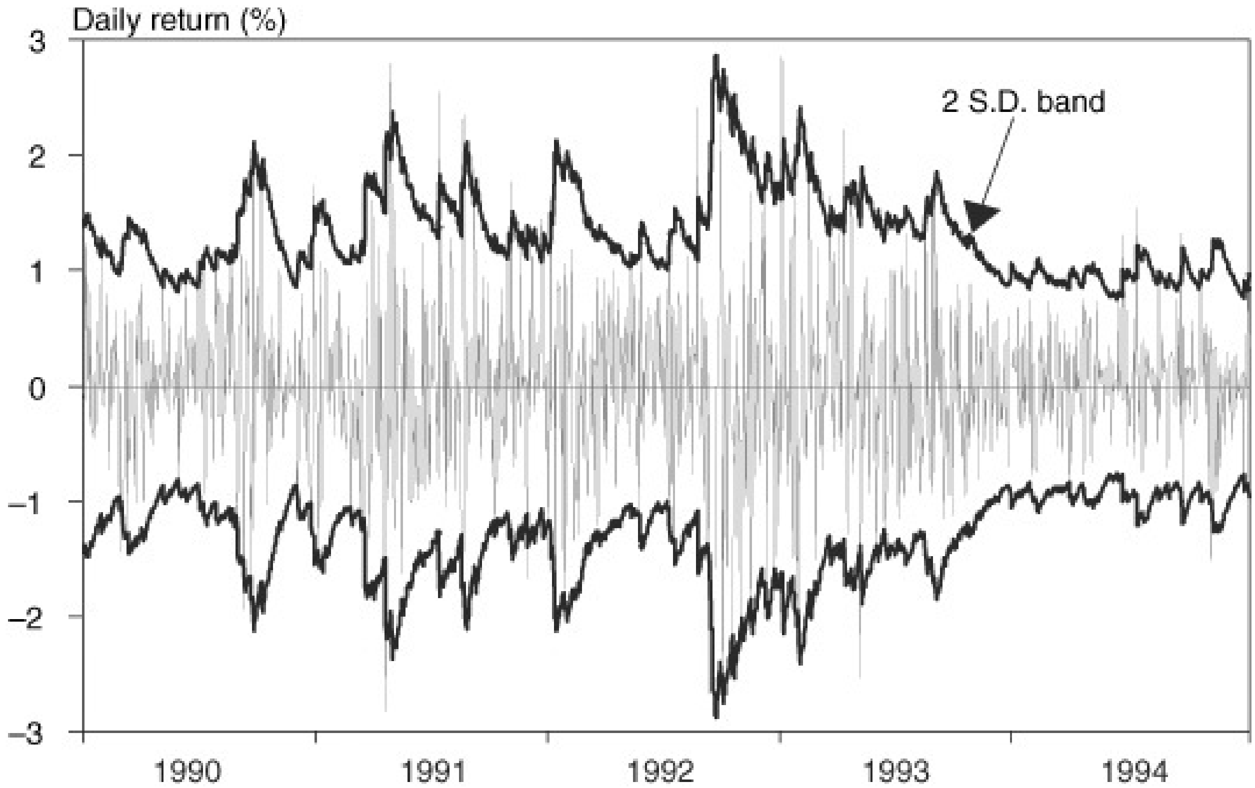
Figure 9-4 displays the GARCH forecast of volatility for the $/BP rate. It shows increased volatility in the fall of 1992. Afterward, volatility decreases progressively over time, not in the abrupt fashion observed in Figure 9-3.

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 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.

FIGURE 9-5
Returns and GARCH confidence bands.



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,

$$r_{t,T} = r_t + r_{t+1} + r_{t+2} + \dots + r_T$$

Let us define n as the number of days, or $T–t + 1 = n$.

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

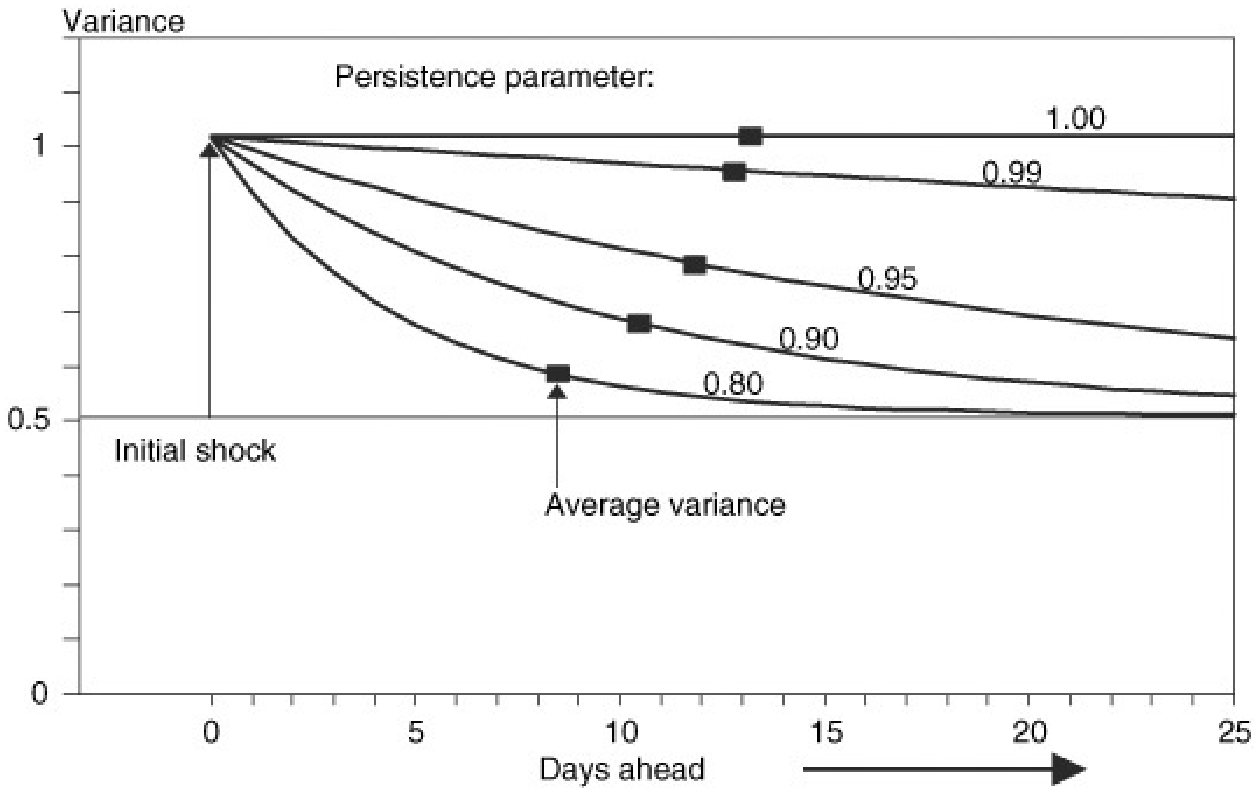
$$E_{t-1}(r_{t,T}^2) = hn \tag{9.8}$$

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 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.

FIGURE 9-6

Mean reversion for the variance.



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.

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 – λ), that is,

$$h_t = \lambda h_{t-1} + (1-\lambda)\, r_{t-1}^2 \qquad (9.9)$$

Here, the λ parameter is called the *decay factor* and must be less than unity.

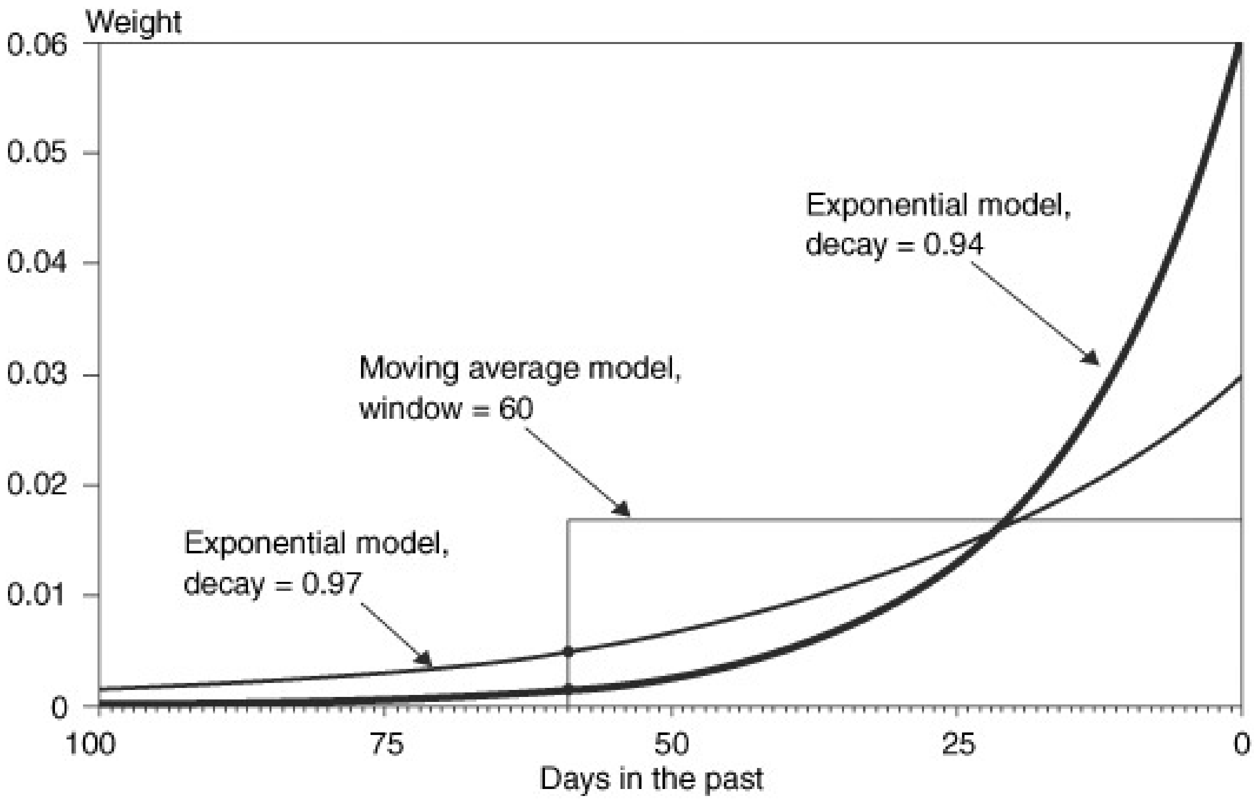
The exponential model places geometrically declining weights on past observations, thus assigning greater importance to recent observations. By recursively replacing h_{t-1} in Equation (9.9), we can write

$$h_t = (1 - \lambda)\,(r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + \cdots) \qquad (9.10)$$

Figure 9-7 displays the pattern of weights for λ = 0.94 and λ = 0.97. For λ = 0.94, the most recent weight is 1 – 0.94 = 0.06. After that, the weights decay fairly quickly, dropping below 0.00012 for data more than 100 days old. Thus the number of *effective* observations is rather small.

FIGURE 9-7

Weights on past observations.



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.

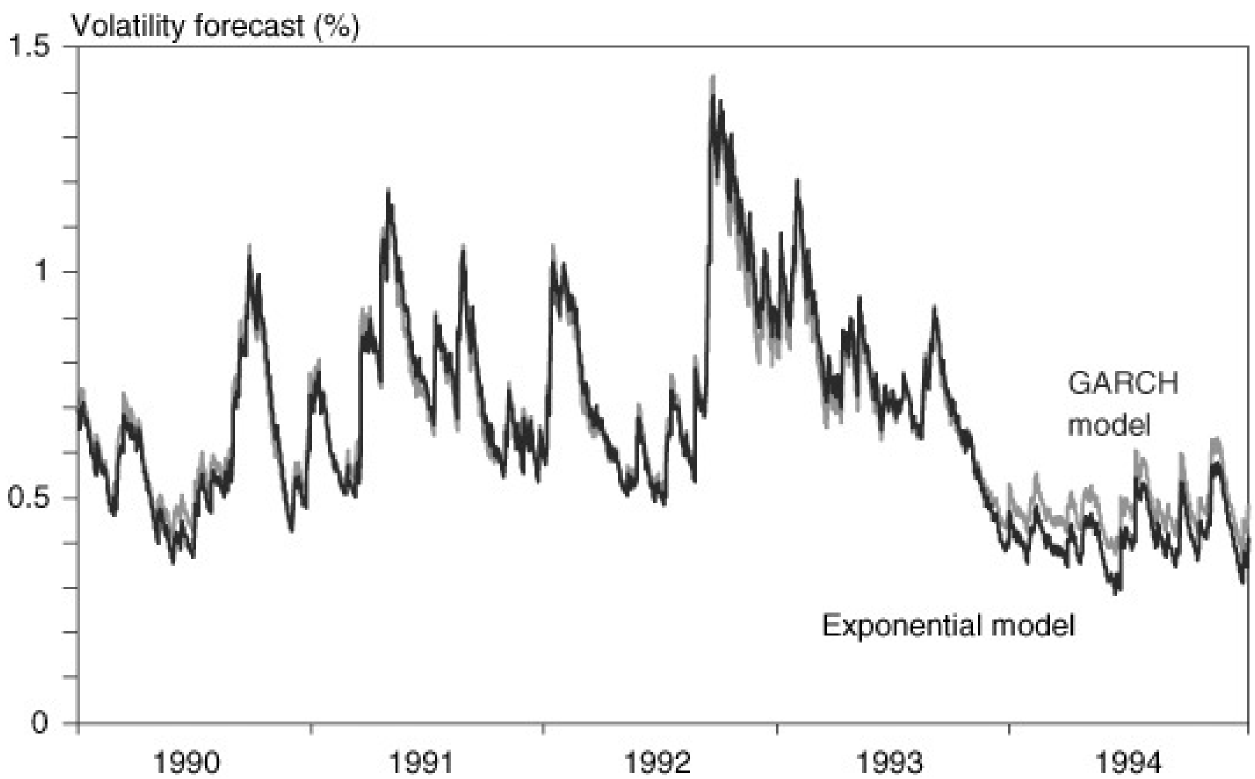
The exponential model is particularly easy to implement because it relies on one parameter only. Thus it is more robust to estimation error than other models. In addition, as was the case for the GARCH model, the estimator is *recursive*;

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 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.

FIGURE 9-8

Exponential volatility forecast.



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.

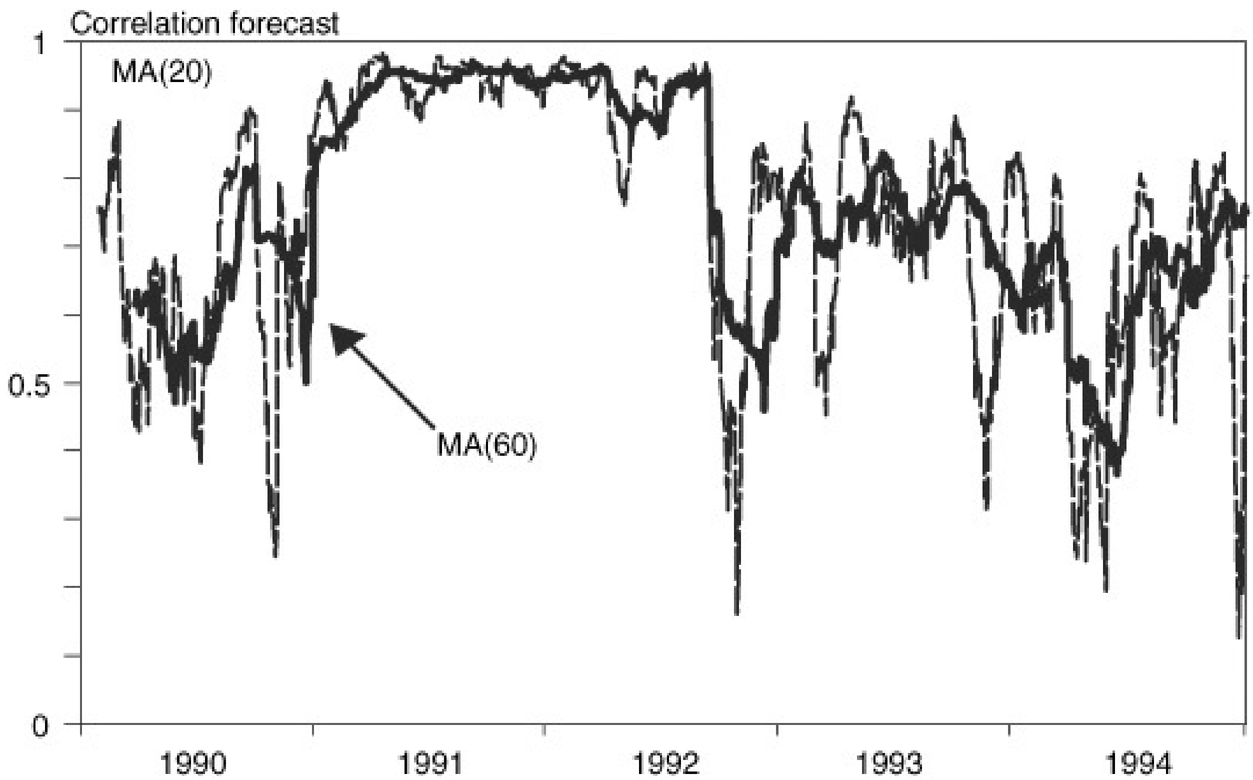
One advantage of multivariate volatility models is that they provide internally consistent risk estimates for a portfolio of assets. Another approach would be to construct the portfolio return series for given weights and to fit a univariate GARCH model to this aggregate series. If the weights change, 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 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.

FIGURE 9-9

Moving-average correlation: $/BP and $/DM.



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 and so on for $h_{12,t}$, the covariance term, and $h_{22,t}$, the second variance term.

$$h_{11,t} = \alpha_{0,11} + \alpha_{1,11} r_{1,t-1}^2 + \alpha_{1,12} r_{1,t-1} r_{2,t-1} + \alpha_{1,13} r_{2,t-1}^2 + \beta_{11} h_{11,t-1} + \beta_{12} h_{12,t-1} + \beta_{13} h_{22,t-1} \tag{9.12}$$

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.

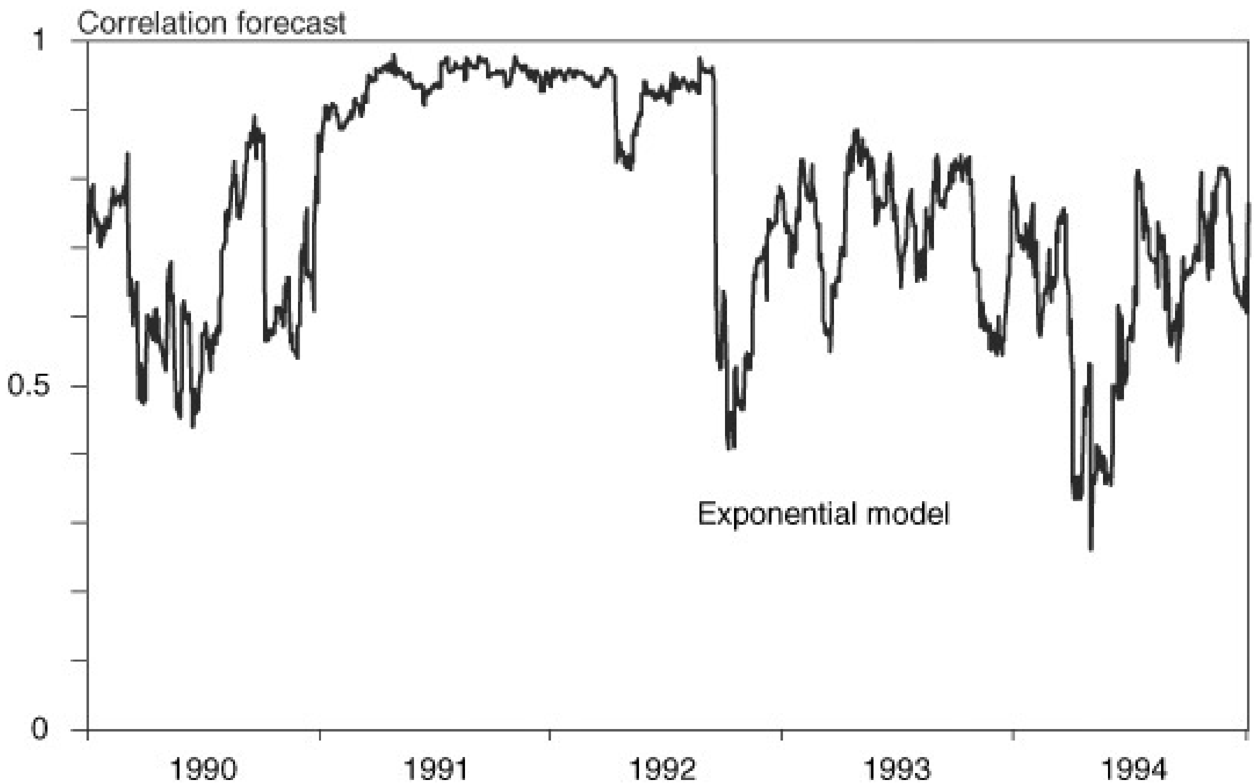
Note that the reason why RiskMetrics sets a common factor λ across all series is to ensure that all estimates of ρ are between –1 and 1. Otherwise, there is no guarantee that this will always be the case.

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, where f represents, for instance, the Black-Scholes function for European options.

$$c_{\text{market}} = f(\sigma_{\text{ISD}}) \tag{9.15}$$

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 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

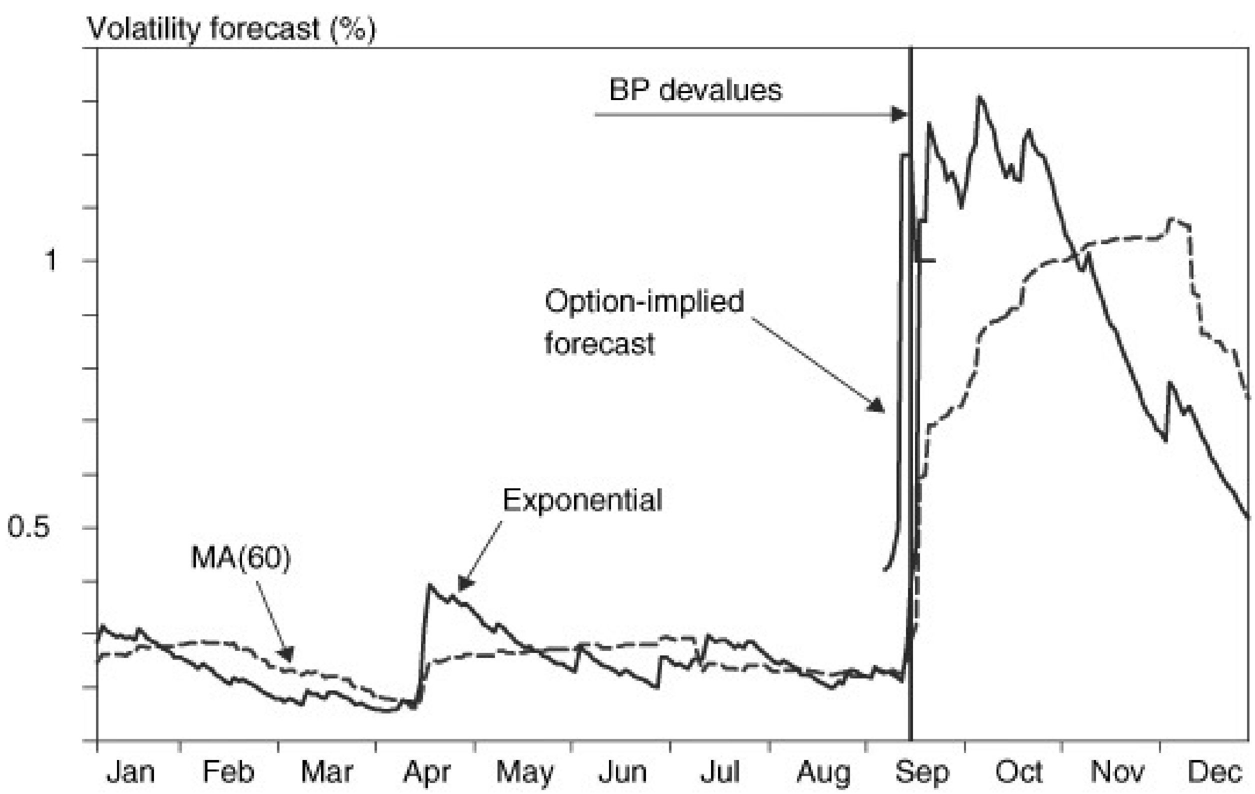
If options markets are efficient, the ISD should provide the market's best estimate of future volatility. After all, options trading involves taking volatility bets. Expressing a view on volatility has become so pervasive in the options markets that prices are often quoted in terms of bid-ask volatility. 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.

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

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} r^2_{1,t-1} \\ r_{1,t-1}r_{2,t-1} \\ r^2_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix} \quad (9.16)$$

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 X 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 X 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 $\in_t$ 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.

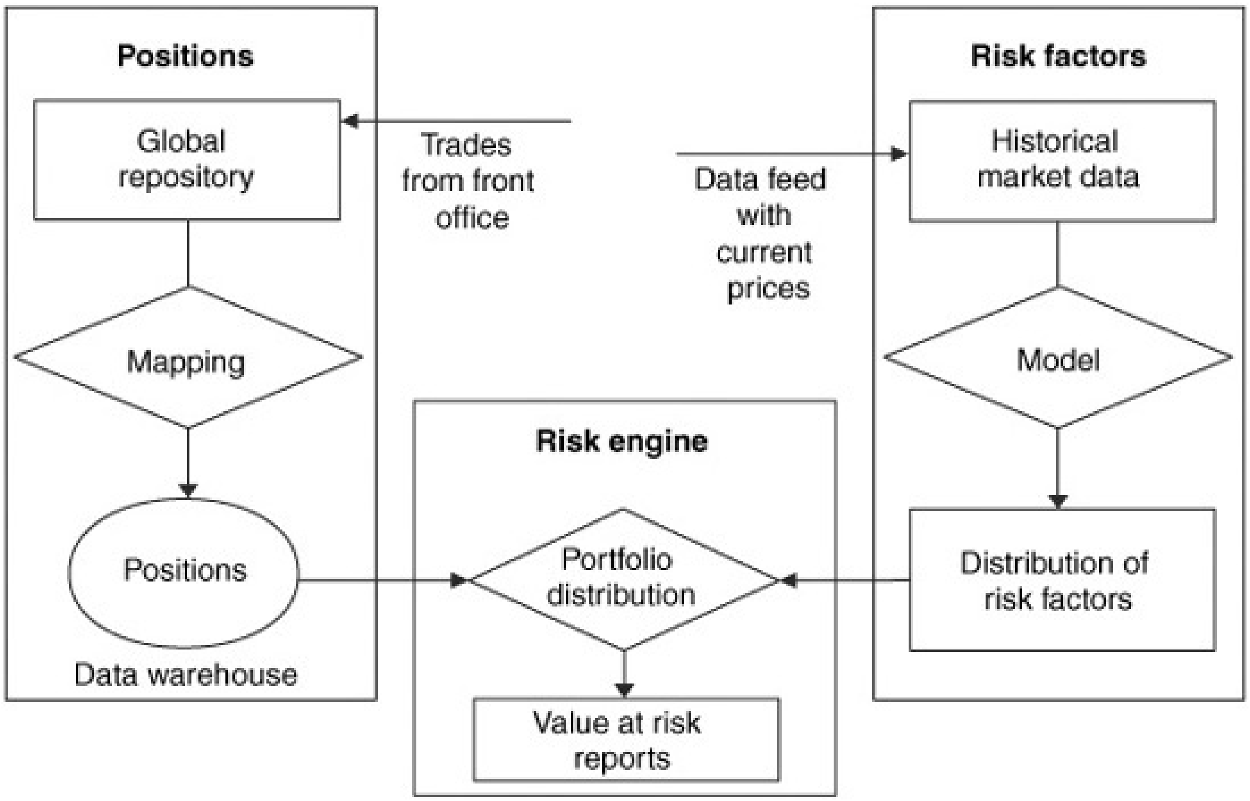
This dichotomy is reflected in the structure of risk management systems, which is described in Figure 10-1. The left-hand side describes the portfolio *positions*, which have as input trades from the front office. The right-hand side describes the *risk factors*, which have as input data feeds with current 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 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.

FIGURE 10-1
VAR systems.



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 which involves the potential change in prices dS. Here, the dollar exposure is given by $x = \Delta_0 S$.

$$dV = \frac{\partial V}{\partial S}\Big|_0 \, dS = \Delta_0 \times dS = (\Delta_0 S)\frac{dS}{S} \tag{10.2}$$

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, 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.

$$\text{VAR} = |\Delta_0| \times \text{VAR}_S = |\Delta_0| \times (\alpha\sigma S_0) \tag{10.3}$$

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

$$dV = (-D^*V)dy \tag{10.4}$$

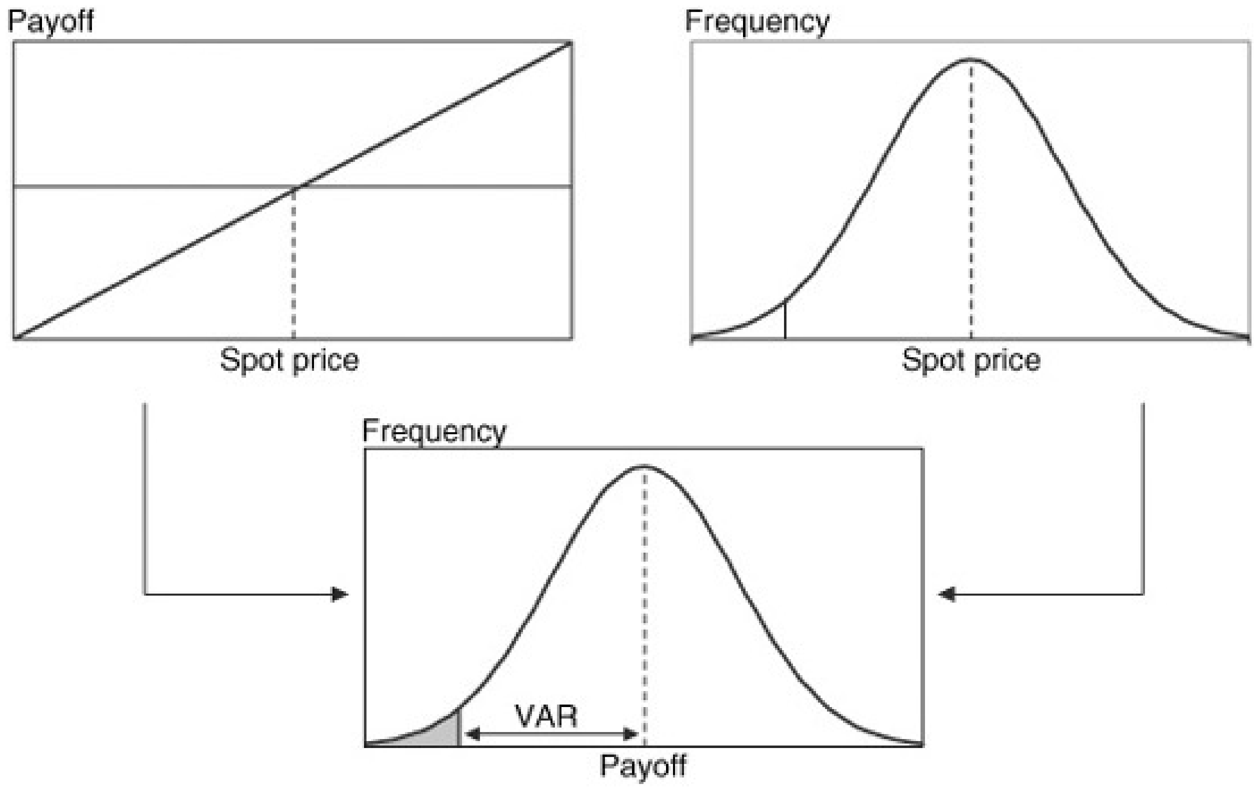
where D^* is the *modified duration*. Here, the dollar exposure is given by $x = -D^*V$. In this case, the portfolio VAR is

$$\text{VAR} = |D^*V| \times (\alpha\sigma) \tag{10.5}$$

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.

FIGURE 10-2

Distribution with linear exposures.



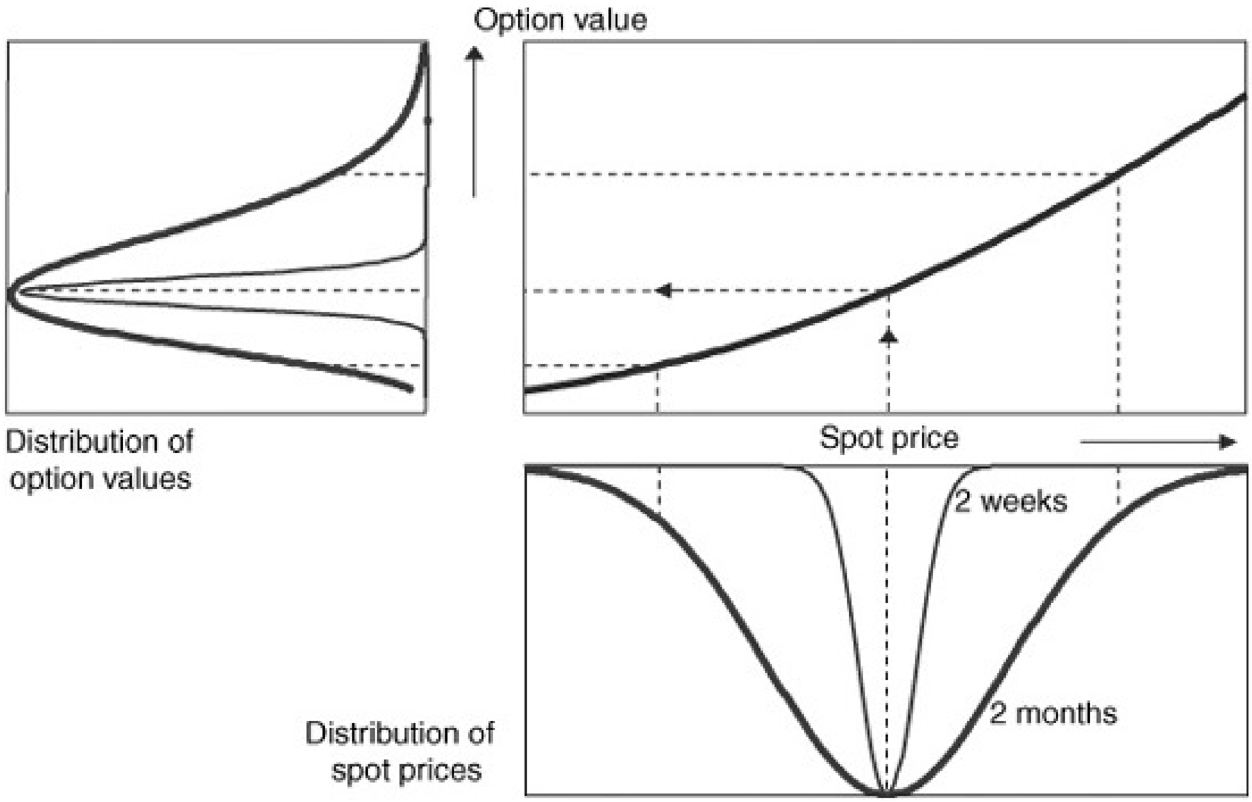
This method is illustrated in Figure 10-2, where the profit payoff *V* is a linear function of the underlying spot price *S* and is displayed in the upper left panel. As shown in the right panel, the price is normally distributed. As a result, the profit itself is normally distributed, as shown at the 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 *c*th quantile for *V* is simply the function evaluated at the *c*th 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

$$\mathrm{VAR} = V(S_0) - V(S_0 - \alpha\sigma S_0) \qquad (10.6)$$

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,

$$\frac{dS}{S} \approx N(0, \sigma^2) \tag{10.8}$$

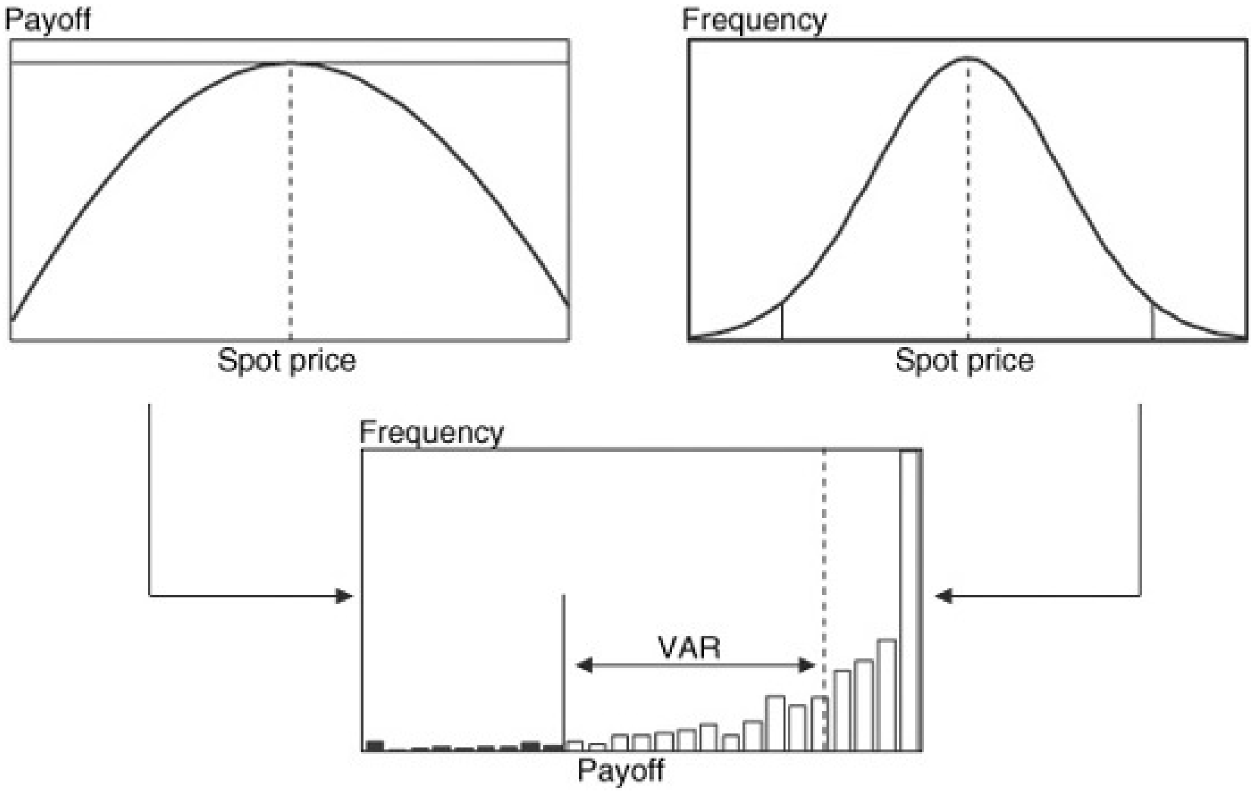
Alternatively, the *historical simulation approach* simply samples from recent historical data.

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 timedecay 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.

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,

$$dV = \frac{\partial V}{\partial S} dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2} dS^2 + \frac{\partial V}{\partial t} dt + \cdots = \Delta dS + \frac{1}{2}\Gamma dS^2 + \Theta dt + \cdots \qquad (10.9)$$

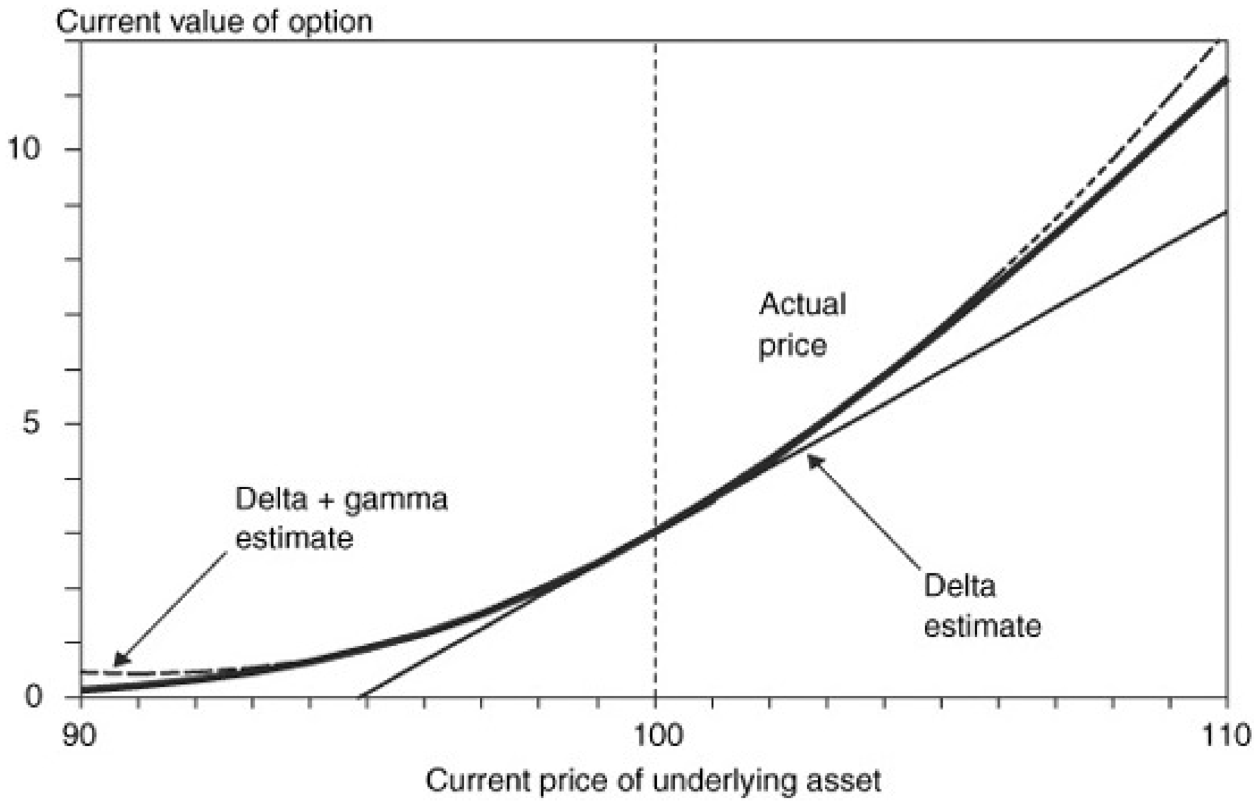
where Γ is now the second derivative of the portfolio value, and Θ is the time drift, which is deterministic. For a fixed-income portfolio, the instantaneous price-yield relationship is now where the second-order coefficient C is called *convexity* and is akin to Γ.

$$dV = -(D^*V)dy + \tfrac{1}{2}(CV)dy^2 + \cdots \qquad (10.10)$$

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 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.

FIGURE 10-5

Delta-gamma approximation for a long call.



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

$$\begin{aligned}
\mathrm{VAR} &= V(S_0) - V(S_0 - \alpha\sigma S_0) \\
&= V(S_0) - [V(S_0) + \Delta(-\alpha\sigma S) + \tfrac{1}{2}\Gamma(-\alpha\sigma S)^2] \qquad (10.11) \\
&= |\Delta|(\alpha\sigma S) - \tfrac{1}{2}\Gamma(\alpha\sigma S)^2
\end{aligned}$$

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 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.

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

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.

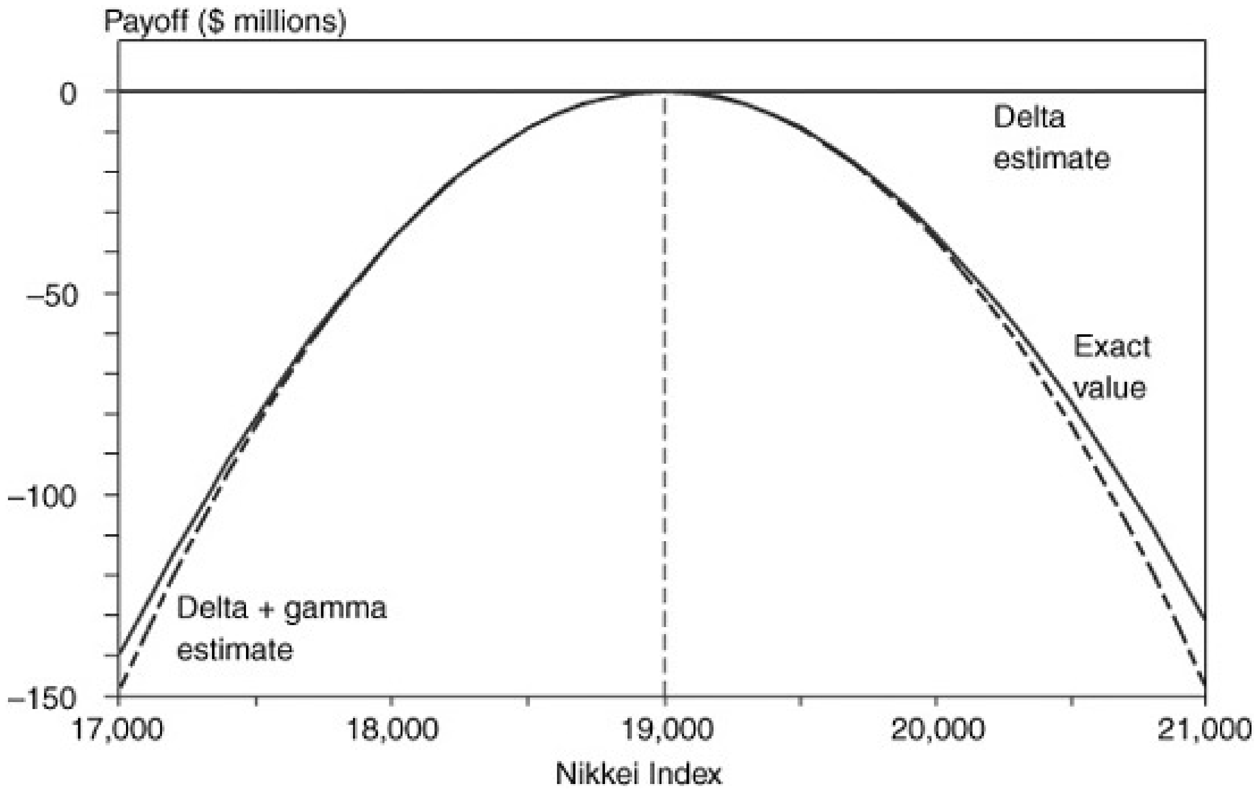
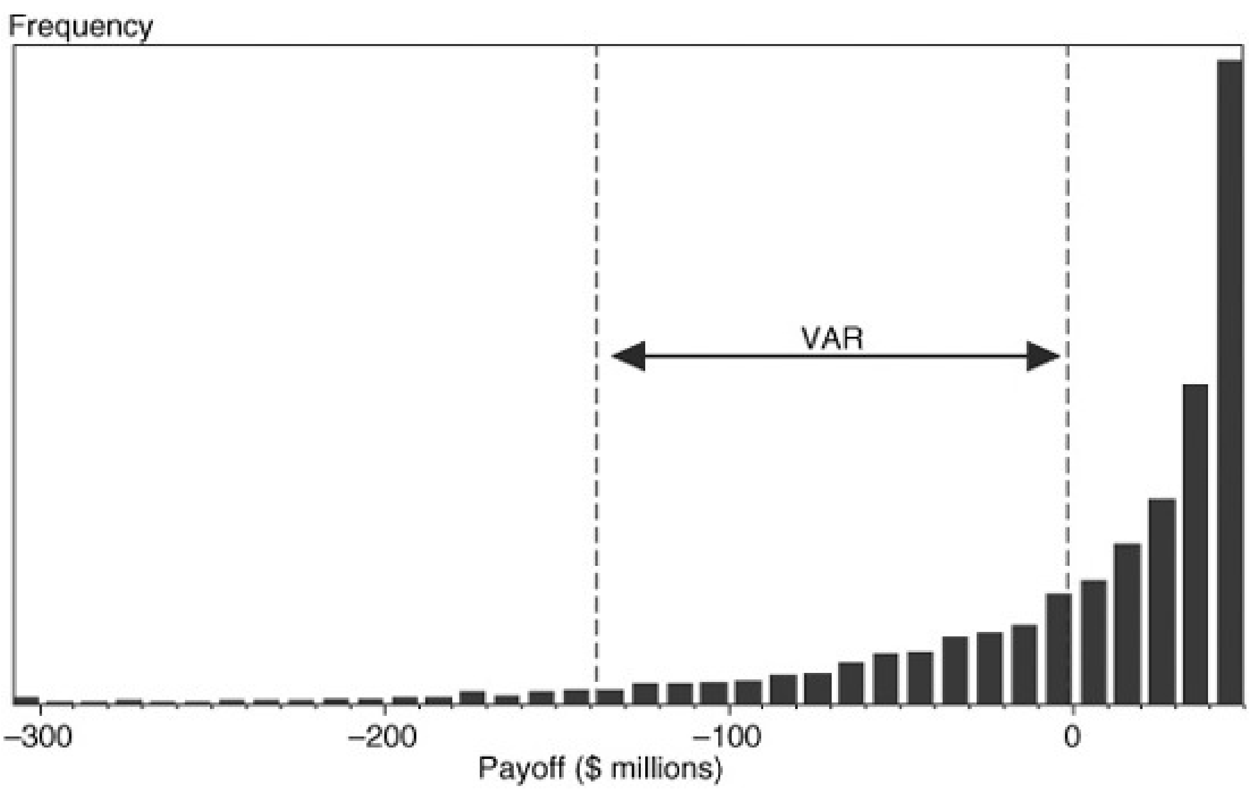


FIGURE 10-7

Distribution of 1-month payoff for straddle.



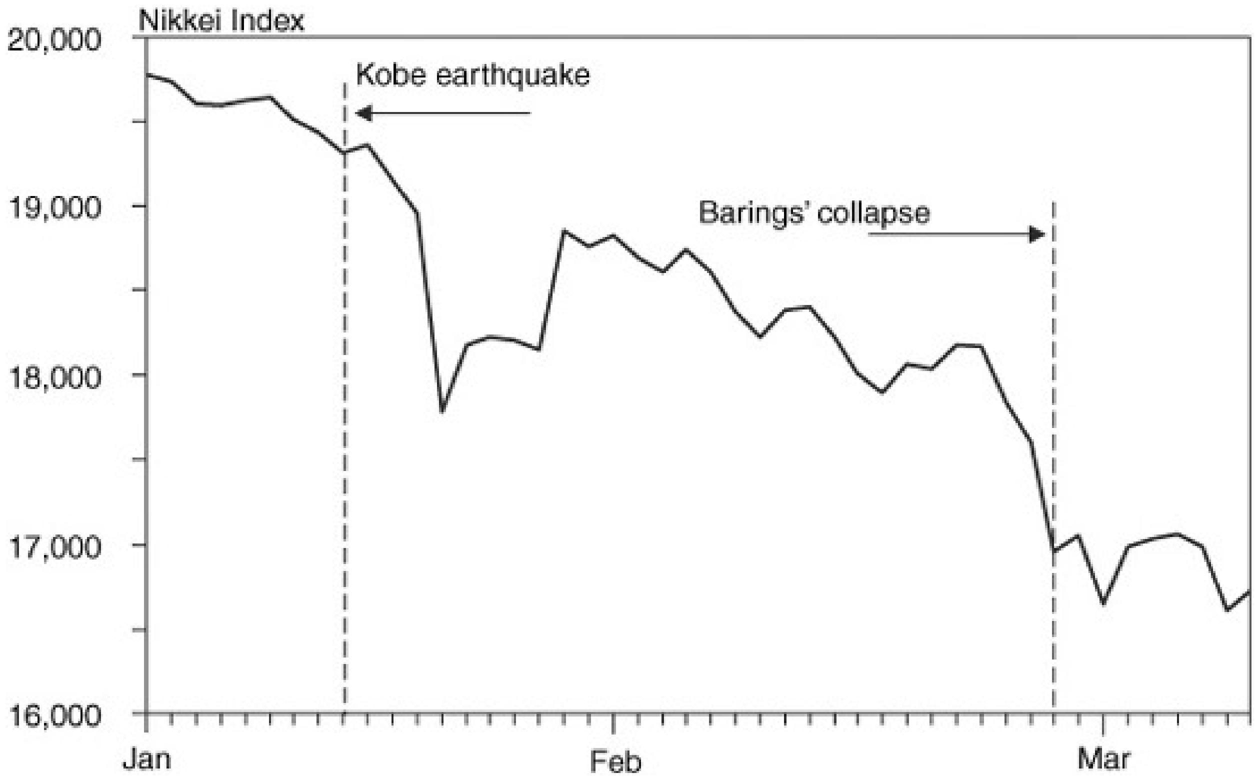
The risks involved are described in Figure 10-7, which plots the frequency distribution of payoffs on the straddle using a *full Monte Carlo simulation* with 10,000 replications. This distribution is obtained from a revaluation of the portfolio after a month over a range of values for the Nikkei. Each 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 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.

FIGURE 10-8

The Nikkei's fall.



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 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.

$$R_{p,t+1} = \sum_{i=1}^{N} w_{i,t} R_{i,t+1} \quad (10.12)$$

Using matrix notations, as in Chapter 7, the portfolio variance is given by

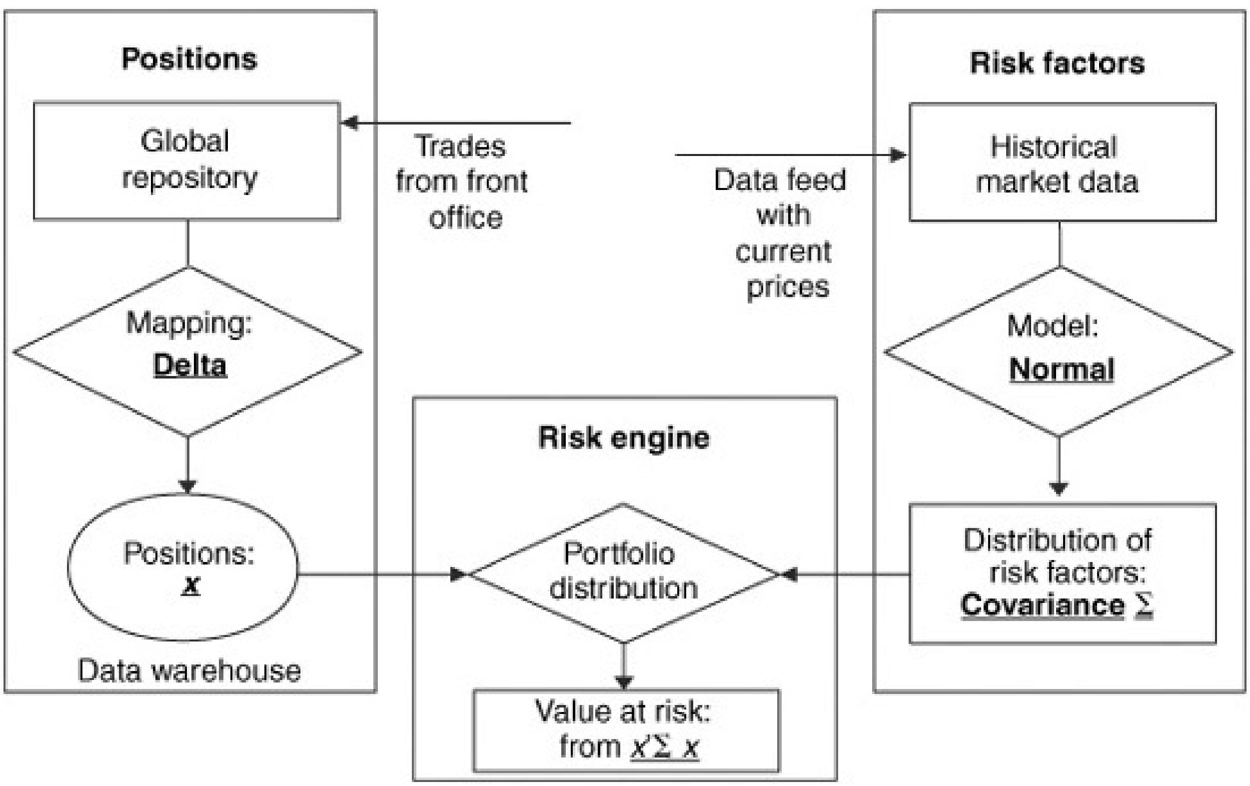
$$\sigma^2(R_{p,t+1}) = w_t' \Sigma_{t+1} w_t \quad (10.13)$$

where $\sum_{t+1}$ is the forecast of the covariance matrix over the VAR horizon, perhaps using models developed in Chapter 9. The portfolio VAR then is 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.

$$\text{VAR} = \alpha\sqrt{x_t' \, \Sigma_{t+1} \, x_t} = \alpha W \sqrt{w_t' \, \Sigma_{t+1} \, w_t} \quad (10.14)$$

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*.

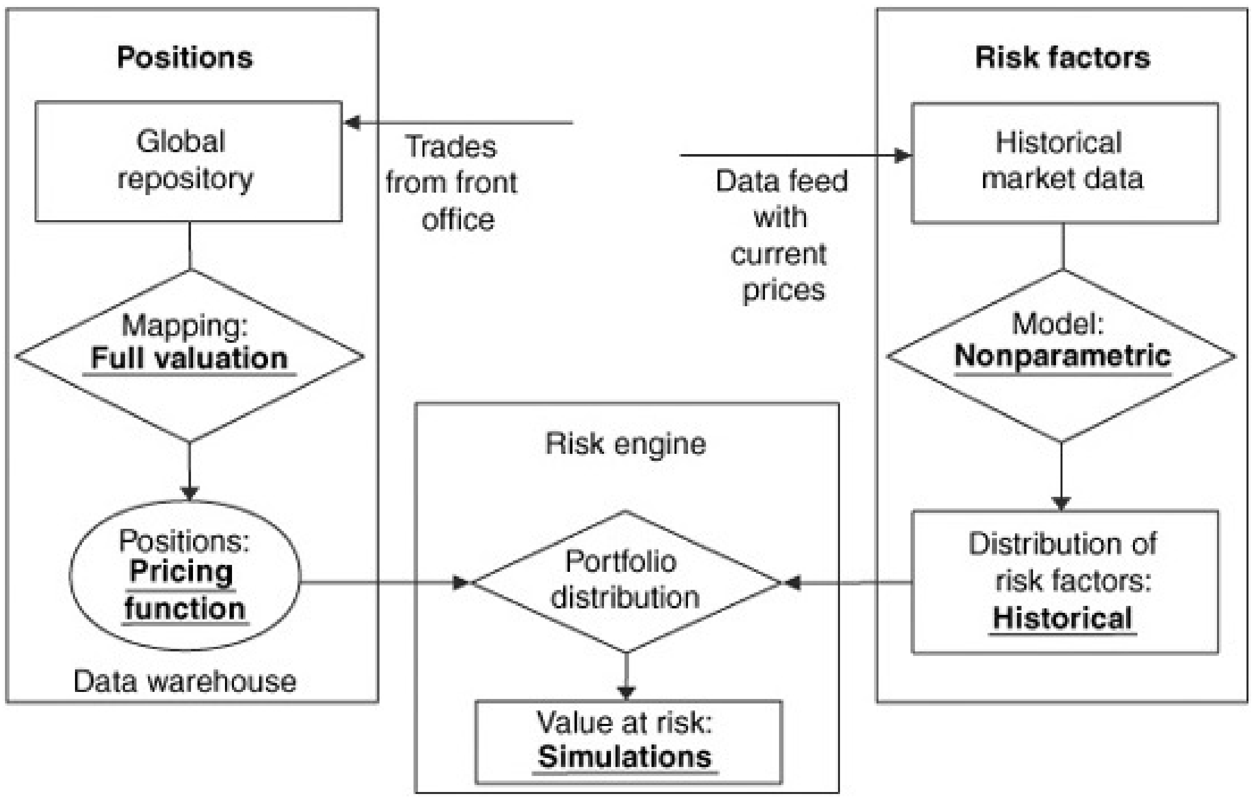
10.4.2 Advantages

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

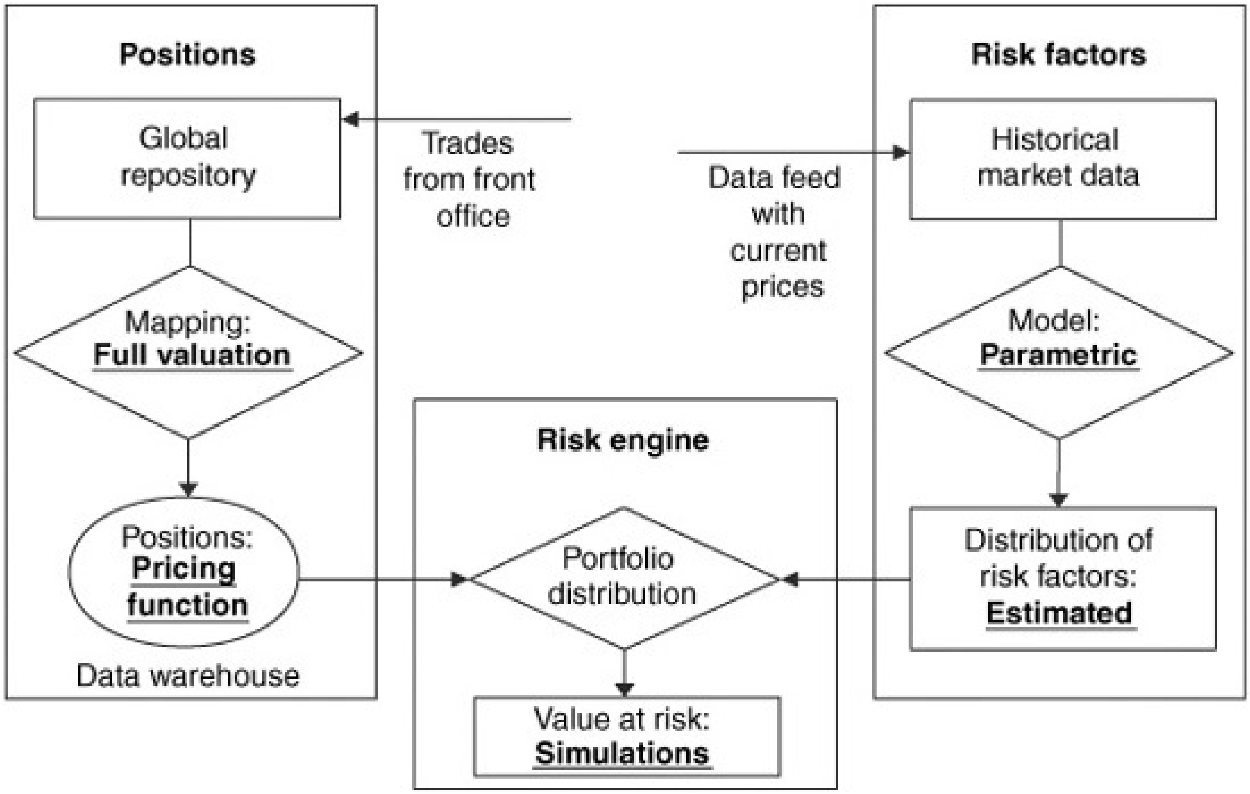
10.5.1 Implementation

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.

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 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.

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

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,

$$dV = \Delta dS + \tfrac{1}{2}\Gamma dS^2 + \cdots \qquad (10.18)$$

In a multivariate framework, the Taylor expansion is 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.

$$dV(S) = \Delta' dS + \tfrac{1}{2}(dS)'\Gamma(dS) + \cdots \qquad (10.19)$$

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

$$\sigma^2(dV) = \Delta^2\sigma^2(dS) + (\tfrac{1}{2}\Gamma)^2\sigma^2(dS^2) + 2(\Delta\tfrac{1}{2}\Gamma)\mathrm{cov}(dS, dS^2) \qquad (10.20)$$

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

$$\sigma^2(dV) = \Delta^2\sigma^2(dS) + \tfrac{1}{2}[\Gamma\sigma^2(dS)]^2 \qquad (10.21)$$

Assume now that the variables dS and dS^2 are jointly normally distributed. Then dV is normally distributed, with VAR given by

$$\mathrm{VAR} = \alpha\sqrt{(\Delta S\sigma)^2 + \tfrac{1}{2}(\Gamma S^2\sigma^2)^2} \qquad (10.22)$$

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 X 0.000422 = $0.0000739 million. Over a 1-month horizon, the standard deviation of the Nikkei is σS = 19,000 X 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 X $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

$$\Psi(t) = E(e^{itX}) \qquad (10.24)$$

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 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.

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)

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.

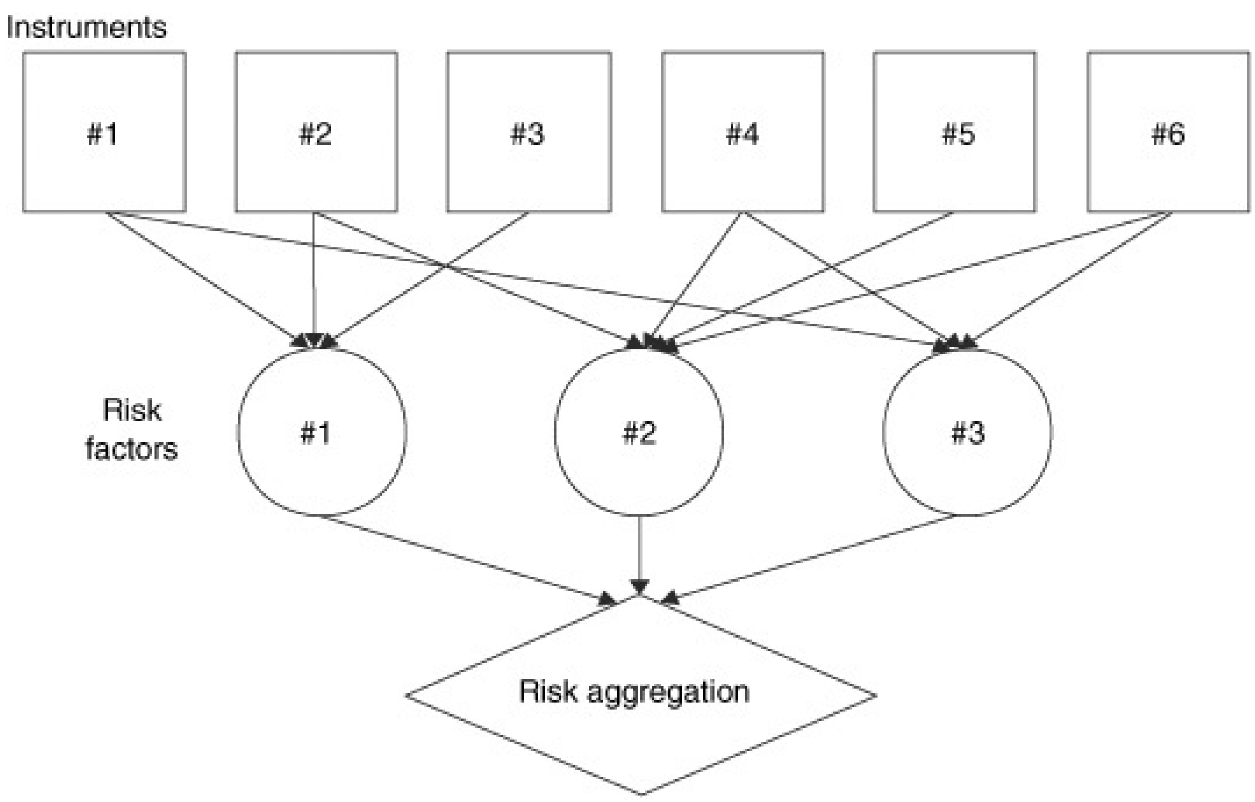


TABLE 11-1
Mapping Exposures

		Exposure on Risk Factor		
	Market Value	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