CHAPTER 5

Computing Value at Risk

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

J.P. Morgan 1994 Annual Report

Perhaps the greatest advantage of value at risk (VAR) is that it summarizes in a single, easy to understand number the downside risk of an institution due to financial market variables. 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, for example, was one of the first users of 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. Shareholders can then 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 portfolio is "frozen" over the horizon or, more generally, that the risk profile of the institution remains constant. In addition, VAR assumes that the current portfolio will be marked-to-market on the target horizon. Section 5.1 shows how to derive VAR figures from probability distributions. This can be done in two ways, either from considering the actual empirical distribution or by approximating the distribution by a parametric approximation, such as the normal distribution, in which case VAR is derived 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 number. 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. 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. Due to normal sampling variation, there is some inherent imprecision in VAR numbers. Thus, observing changes in VAR numbers for different estimation windows is perfectly normal. Section 5.3 provides a framework for analyzing normal sampling variation in VAR and discusses methods to improve the accuracy of VAR figures. Finally, Section 5.4 provides some concluding thoughts.

5.1 COMPUTING VAR

With all the requisite tools in place, we can now formally define the value at risk (VAR) of a portfolio. *VAR summarizes the expected maximum loss (or worst loss) over a target horizon within a given confidence interval.* Initially, we take the quantitative factors, the horizon and confidence level, as given.

5.1.1 Steps in Constructing VAR

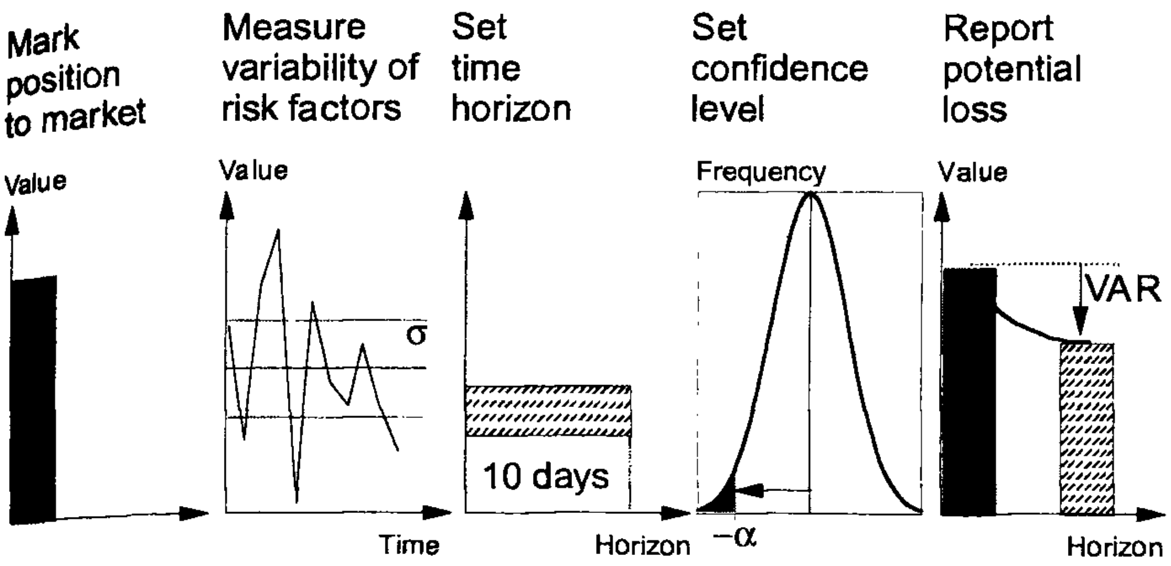
Assume, for instance, that we need to measure the VAR of a $100 million equity portfolio over 10 days at the 99 percent confidence level. The following steps are required to compute VAR:

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

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

FIGURE 5–1

Steps in constructing VAR.



Sample computation:

$100M × 15% × √(10/252) × 2.33 = $7M

5.1.2 VAR for General Distributions

To compute the VAR of a portfolio, define W_0 as the initial investment and R as its rate of return. The portfolio value at the end of the target horizon is $W = W_0\,(1 + R)$. As before, the expected return and volatility of R are μ and σ. Define now the lowest portfolio value at the given confidence level c as $W^* = W_0\,(1 + R^*)$. The *relative VAR* is defined as the dollar loss relative to the mean:

$$\text{VAR(mean)} = E(W) - W^* = -W_0\,(R^* - \mu) \tag{5.1}$$

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

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

In both cases, finding VAR is equivalent to identifying the minimum value W^* or the cutoff return R^*.

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. Its only drawback is that the mean return is sometimes difficult to estimate.

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:

$$c = \int_{W^*}^{\infty} f(w)\, dw \tag{5.3}$$

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

$$1 - c = \int_{-\infty}^{W^*} f(w)\, dw = P(w \leq W^*) = p \tag{5.4}$$

In other words, the area from $-\infty$ to W^* must sum to $p = 1 - c$, for instance, 5 percent. 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. Figure 5–2, for instance, reports J.P. Morgan's distribution of daily revenues in 1994.

To compute VAR, assume that daily revenues are identically and independently distributed. We can then derive the VAR at the 95 percent confidence level from the 5 percent left-side "losing tail" from the histogram.

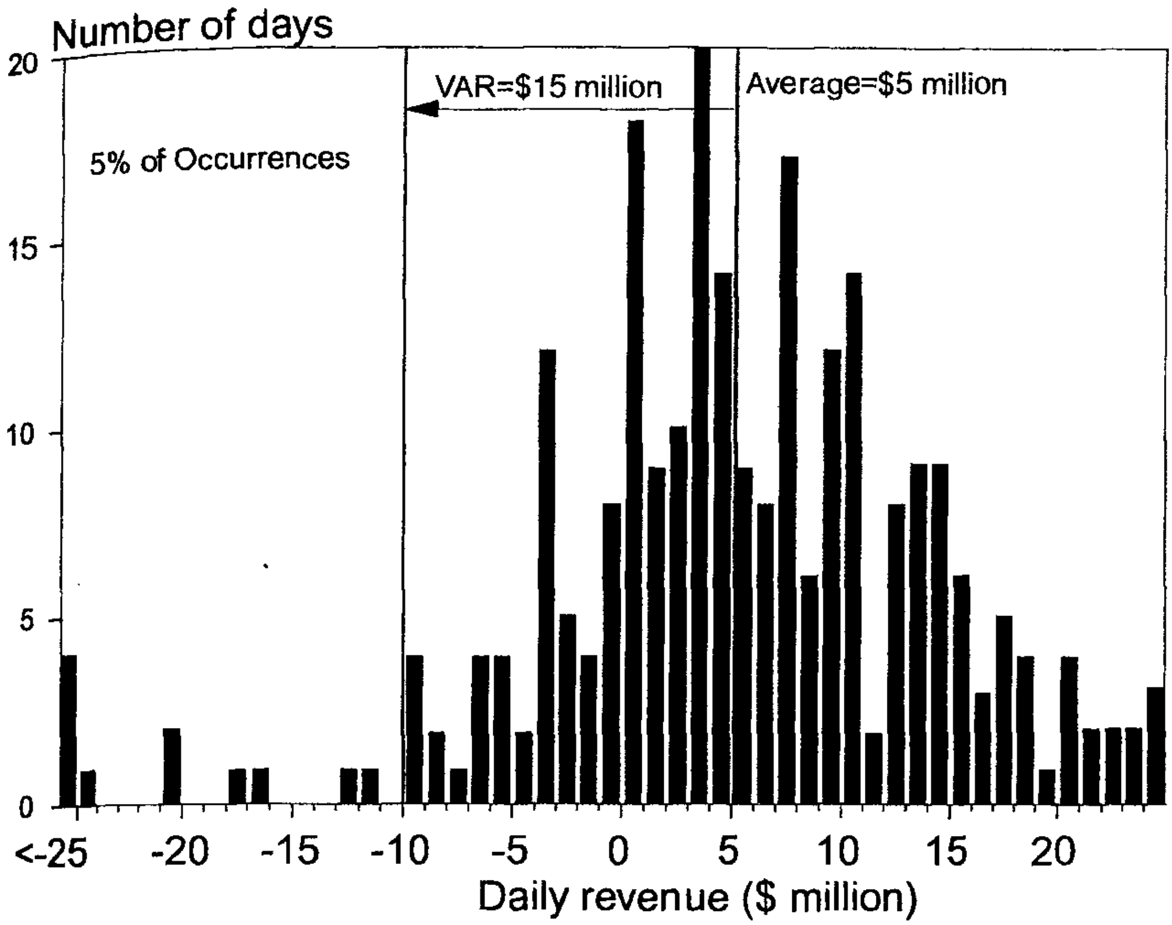
From this graph, the average revenue is about \$5.1 million. There is a total of 254 observations; therefore, we would like to find W^* such that the number of observations to its left is 254×5 percent $= 12.7$. We have 11 observations to the left of $-\$10$ million and 15 to the left of $-\$9$ million. Interpolating, we find $W^* = -\$9.6$ million. The VAR of daily revenues, measured relative to the mean, is VAR $= E(W) - W^* = \$5.1$ million $- (-\$9.6$ million$) = \$14.7$ million. If one wishes to measure VAR in terms of absolute dollar loss, VAR is then \$9.6 million.

5.1.3 VAR for Parametric Distributions

The VAR computation can be simplified considerably if the distribution can be assumed to belong to a parametric family, such as the normal dis-

FIGURE 5–2

Distribution of daily revenues.



tribution. 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 sometimes 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 normal approximation is realistic. If not, another distribution may fit the data better.

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 also can be written as $-|R^*|$.

Further, we can associate R^* with a standard normal deviate $\alpha > 0$ by setting

$$-\alpha = \frac{-|R^*| - \mu}{\sigma} \tag{5.5}$$

It is equivalent to set

$$1 - c = \int_{-\infty}^{W^*} f(w)\, dw = \int_{-\infty}^{-|R^*|} f(r)\, dr = \int_{-\infty}^{-\alpha} \Phi(\epsilon)\, d\epsilon \tag{5.6}$$

Thus the problem of finding a VAR is equivalent to finding the deviate α such that the area to the left of it is equal to $1 - c$. This is made possible by turning to tables of the *cumulative standard normal distribution function,* which is the area to the left of a standard normal variable with value equal to d:

$$N(d) = \int_{-\infty}^{d} \Phi(\epsilon)\, d\epsilon \tag{5.7}$$

This function also plays a key role in the Black-Scholes option pricing model. Figure 5–3 graphs the cumulative density function $N(d)$, which increases monotonically from 0 (for $d = -\infty$) to 1 (for $d = +\infty$), going through 0.5 as d passes through 0.

To find the VAR of a standard normal variable, select the desired left-tail confidence level on the vertical axis, say, 5 percent. This corresponds to a value of $\alpha = 1.65$ below 0. We then retrace our steps, back from the α we just found to the cutoff return R^* and VAR. From Equation (5.5), the cutoff return is

$$R^* = -\alpha\sigma + \mu \tag{5.8}$$

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.

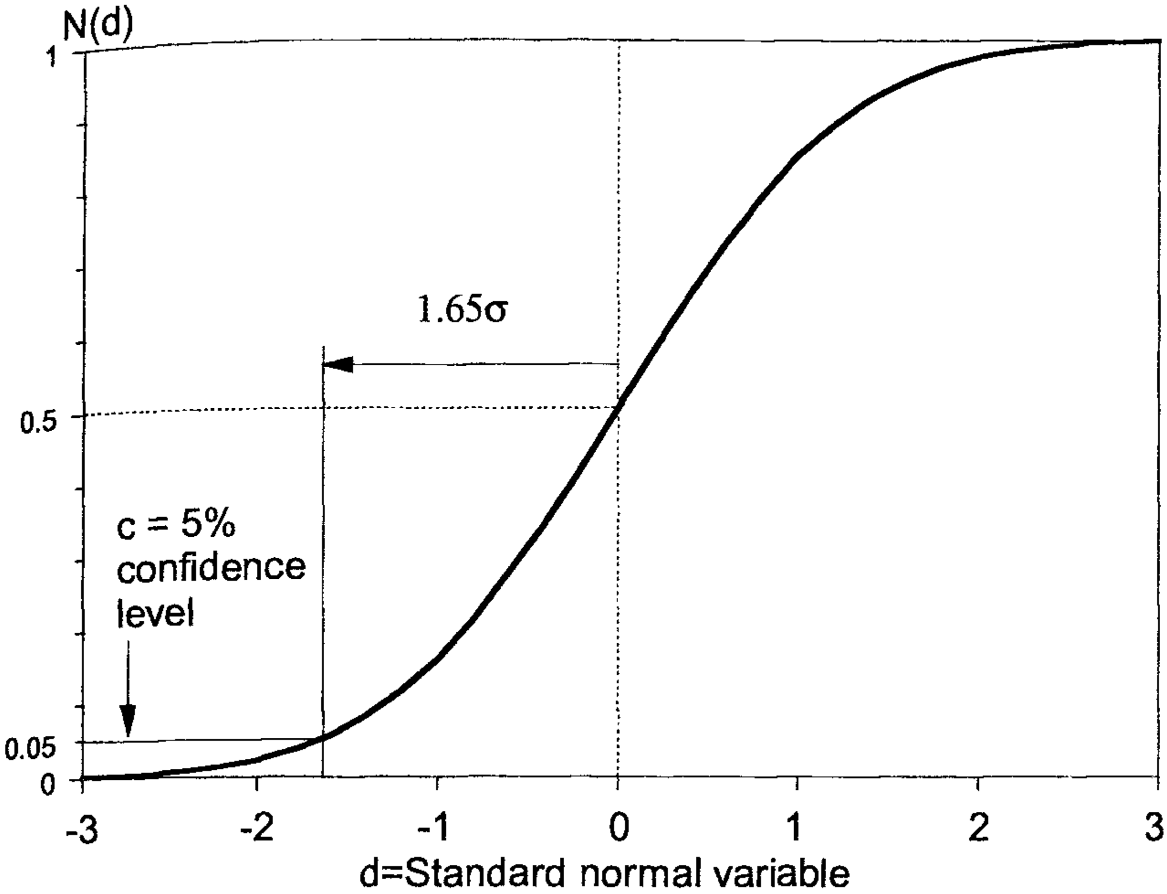
Using Equation (5.1), we find the VAR below the mean as

$$\text{VAR(mean)} = -W_0(R^* - \mu) = W_0\alpha\sigma\sqrt{\Delta t} \tag{5.9}$$

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

FIGURE 5–3

Cumulative normal probability distribution.



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

$$\text{VAR(zero)} = -W_0R^* = W_0(\alpha\sigma\sqrt{\Delta t} - \mu\Delta t) \qquad (5.10)$$

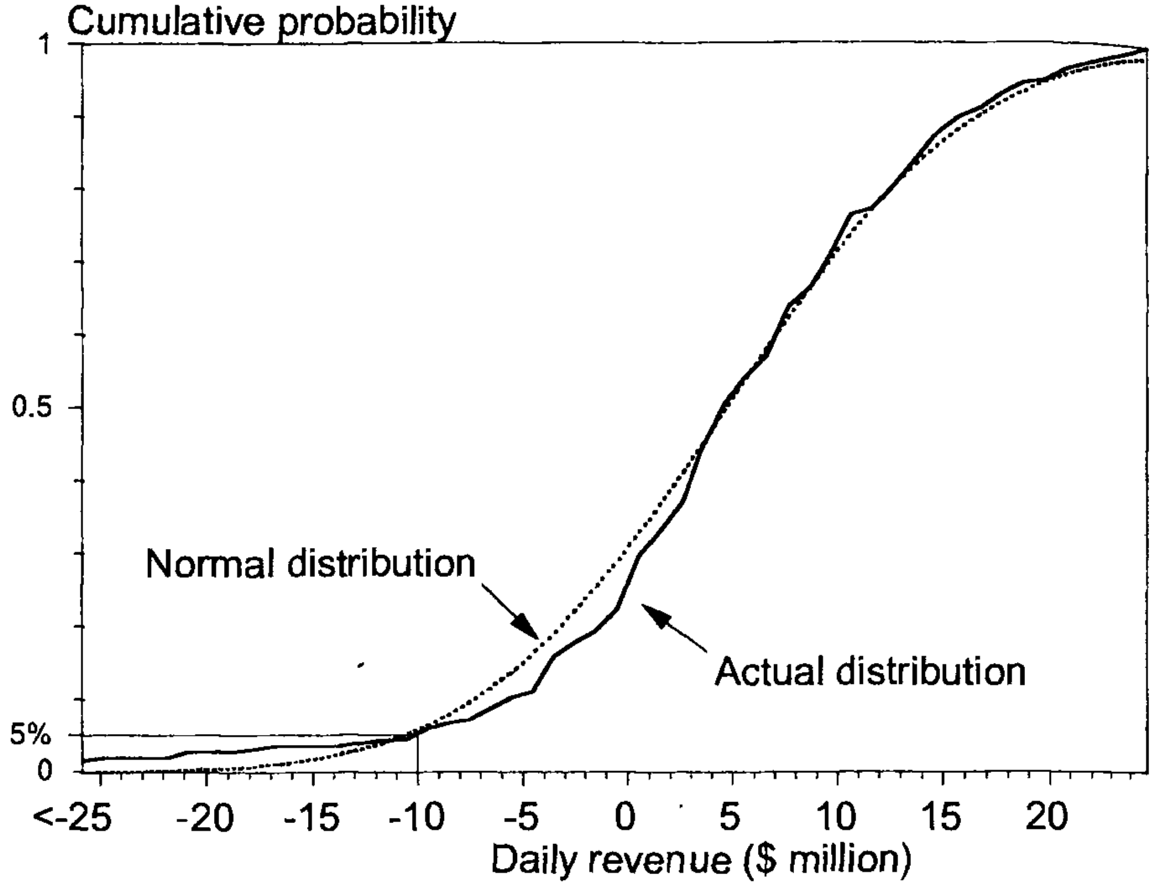
This method generalizes to other cumulative probability functions (cdf) as well as the normal, as long as all the uncertainty is contained in σ. Other distributions will entail different values of α. The normal distribution is just particularly easy to deal with because it adequately represents many empirical distributions. This is especially true for large, well-diversified portfolios but certainly not for portfolios with heavy option components and exposures to a small number of financial risks.

5.1.4 Comparison of Approaches

How well does this approximation work? For some distributions, the fit can be quite good. Consider, for instance, the daily revenues in Figure 5–2. The standard deviation of the distribution is \$9.2 million. According

FIGURE 5–4

Comparison of cumulative distributions.



to Equation (5.9), the normal-distribution VAR is $\alpha \times (\sigma W_0) = 1.65 \times$ \$9.2 million = \$15.2 million. Note that this number is very close to the VAR obtained from the general distribution, which was \$14.7 million.

Indeed, Figure 5–4 presents the cumulative distribution functions (cdf) obtained from the histogram in Figure 5–2 and from its normal approximation. The actual cdf is obtained from summing, starting from the left, all numbers of occurrences in Figure 5–2 and then scaling by the total number of observations. The normal cdf is the same as that in Figure 5–3, with the horizontal axis scaled back into dollar revenues using Equation (5.8). The two lines are generally very close, suggesting that the normal approximation provides a good fit to the actual data.

5.1.5 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:

$$L = \alpha\sigma - \mu \tag{5.11}$$

which is an early description of Equation (5.10).

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

More recently, Artzner et al. (1999) list four desirable properties for risk measures for capital adequacy purposes. A risk measure can be viewed as a function of the distribution of portfolio value W, which is summarized into a single number $\rho(W)$:

- *Monotonicity:* If $W_1 \leq W_2$, $\rho(W_1) \geq \rho(W_2)$, or if a portfolio has systematically lower returns than another for all states of the world, its risk must be greater.
- *Translation invariance.* $\rho(W + k) = \rho(W) - k$, or adding cash k to a portfolio should reduce its risk by k.
- *Homogeneity.* $\rho(bW) = b\rho(W)$, or 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.* $\rho(W_1 + W_2) \leq \rho(W_1) + \rho(W_2)$, or merging portfolios cannot increase risk.

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. One can also show that the shortfall measure $E(-X|X \leq -VAR)$, which is the expected loss conditional on exceeding VAR, satisfies these desirable "coherence" properties.

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.

Of course, the preceding discussion does not consider another essential component for portfolio comparisons: expected returns. In practice, one obviously would want to balance increasing risk against increasing expected returns. The great benefit of VAR, however, is that it brings attention and transparency to the measure of risk, a component of the decision process that is not intuitive and as a result too often ignored.

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. Bankers Trust, for instance, has long used a 99 percent VAR over an annual horizon to compare the risks of various units. Assuming a normal distribution, we show later that it is easy to convert disparate bank measures into a common number.

The focus here 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 larger bets. For this purpose, the choice of the confidence level and horizon does 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 is generally 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. Higher confidence levels will generate higher VAR figures.

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 VAR confidence level. Table 5–1, for instance, shows that to maintain a Baa investment-grade credit rating, the institution should have a default probability of 0.17 percent over the next year. It therefore should carry enough capital to cover its annual VAR at the 99.83 percent confidence level, or 100 − 0.17 percent.

Longer horizons, with a constant risk profile, inevitably lead to higher default frequencies. Institutions with an initial Baa credit rating have a default frequency of 10.50 percent over the next 10 years. The same credit rating can be achieved by extending the horizon or decreasing the confidence level appropriately. These two factors are intimately related.

TABLE 5–1

Credit Rating and Default Rates

	Default Frequency	
Desired Rating	**1 Year**	**10 Yeara**
Aaa	0.02%	1.49%
Aa	0.05%	3.24%
A	0.09%	5.65%
Baa	0.17%	10.50%
Ba	0.77%	21.24%
B	2.32%	37.98%

Source: Adapted from Moody's default rates from 1920–1998.

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 is described in a later chapter. 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

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 is then 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 quite arbitrary. There are many combinations of the confidence level, the horizon, and the multiplicative factor that would yield the same capital charge. The origin of the factor k also looks rather mysterious.

Presumably, the multiplicative factor 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 due to a short sample period, 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.

Stahl (1997) justifies the choice of k based on Chebyshev's inequality. For any random variable x with finite variance, the probability of falling outside a specified interval is

$$P(|x - \mu| > r\sigma) \leq 1/r^2 \tag{5.12}$$

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

$$P[(x - \mu) < -r\sigma] \leq \tfrac{1}{2}\, 1/r^2 \tag{5.13}$$

We now set the right-hand side of this inequality to the desired level of 1 percent. This yields $r(99\%) = 7.071$. The maximum VAR is therefore $VAR_{\max} = r(99\%)\sigma$.

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

$$\text{VAR}_N = \alpha(99\%)\sigma = 2.326\sigma \tag{5.14}$$

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

$$k = \frac{\text{VAR}_{\text{max}}}{\text{VAR}_N} = \frac{7.071\sigma}{2.326\sigma} = 3.03 \tag{5.15}$$

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

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 measures into the Basel Committee internal models measures. RiskMetrics provides a 95 percent confidence interval (1.65σ) over 1 day. The Basel Committee rules define a 99 percent confidence interval (2.33σ) over 10 days. The adjustment takes the following form:

$$\mathrm{VAR}_{\mathrm{BC}} = \mathrm{VAR}_{\mathrm{RM}} \frac{2.33}{1.65} \sqrt{10} = 4.45\mathrm{VAR}_{\mathrm{RM}}$$

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

More generally, Table 5–2 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 DM/$

TABLE 5–2

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

Confidence Level c	Number of S.D. α	Horizon Δt	Actual S.D. $\sigma\sqrt{\Delta t}$	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

exchange rate (now the euro/$ rate). 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 due to limited sample size. Users should beware of the limited precision behind the reported VAR numbers.

5.3.1 The Problem of Measurement Errors

From the viewpoint of VAR users, it is important 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 management is highly confident in this figure or that it is 95 percent sure that the true estimate is in 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 quite a different picture of VAR. The first is very precise; the second is rather uninformative (although it tells us that it is not in the hundreds of millions of dollars). This is why it is useful to examine measurement errors in VAR figures.

Consider a situation where VAR is obtained from the historical simulation method, which uses a historical window of T days to measure risk. The problem is that the reported VAR measure is only an *estimate* of the true value and is affected by sampling variability. In other words, different choices of the window T will lead to different VAR figures.

One possible interpretation of the estimates (the view of "frequentist" statisticians) is that these estimates $\hat{\mu}$ and $\hat{\sigma}$ are samples from an underlying distribution with unknown parameters μ and σ. 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 series, like emerging markets, 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 parameters $\hat{\mu}$ and $\hat{\sigma}$. We now turn to a description of the distribution of statistics on which VAR measures are based.

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

$$\hat{\mu} \approx N(\mu, \sigma^2/T) \tag{5.16}$$

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 rate of $\sigma\sqrt{1/T}$ as T increases.

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

$$\frac{(T-1)\,\hat{\sigma}^2}{\sigma^2} \approx \chi^2(T-1) \tag{5.17}$$

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

$$\hat{\sigma}^2 \approx N\left(\sigma^2, \sigma^4 \frac{2}{T-1}\right) \tag{5.18}$$

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

$$se(\hat{\sigma}) = \sigma \sqrt{\frac{1}{2T}} \tag{5.19}$$

For instance, consider monthly returns on the DM/$ rate from 1973 to 1998. Sample parameters are $\hat{\mu} = -0.15$ percent, $\hat{\sigma} = 3.39$ percent, with $T = 312$ observations. The standard error of the estimate indicates how confident we are about the sample value; the smaller the error, the more confident we are. One standard error in $\hat{\mu}$ is $se(\hat{\mu}) = \hat{\sigma}\sqrt{1/T} =$ $3.39\sqrt{1/312} = 0.19$ percent. Therefore, the point estimate of $\hat{\mu} = -0.15$ percent is less than one standard error away from 0. Even with 26 years of data, μ is measured very imprecisely.

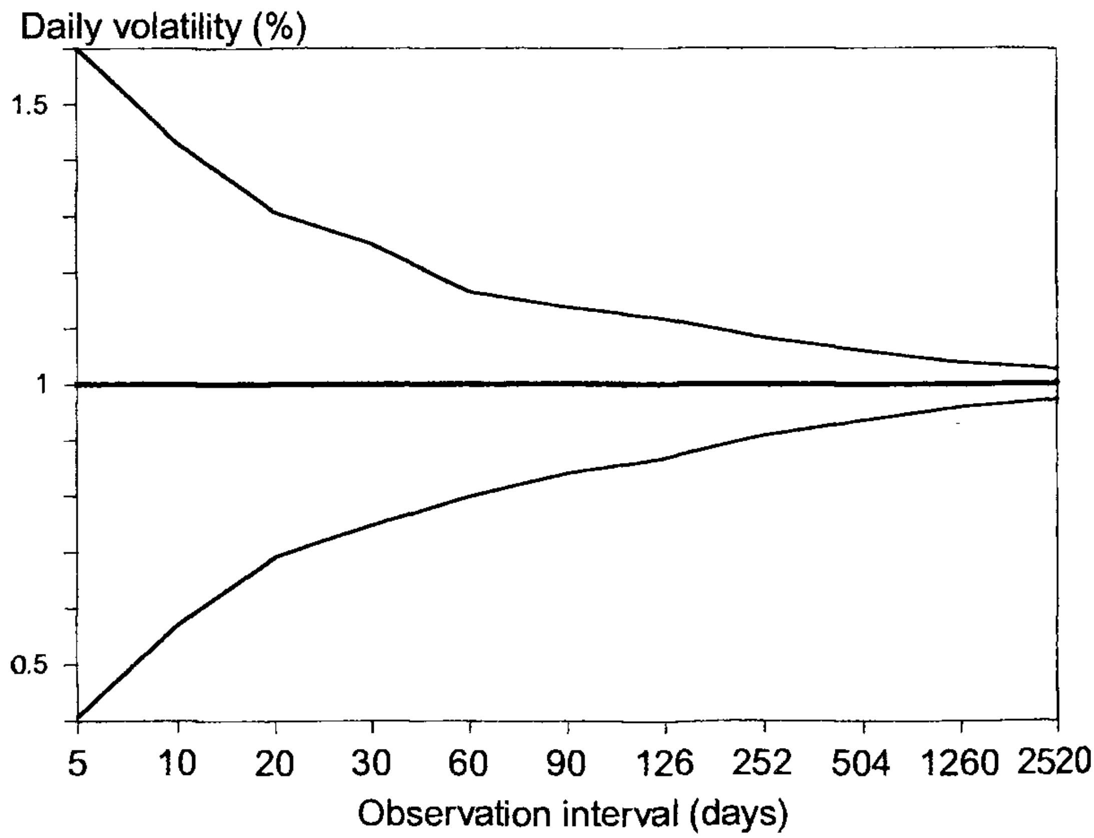
In contrast, one standard error for $\hat{\sigma}$ is $se(\hat{\sigma}) = \hat{\sigma}\sqrt{1/2T} = 3.39\sqrt{1/624} = 0.14$ percent. Since this number is much smaller than the estimate of 3.39 percent, we can conclude that the volatility is estimated with much greater accuracy than the expected return—giving some confidence in the use of VAR systems.

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

With 5 trading days, the band is rather imprecise, with upper and lower values set at [0.41%, 1.60%]. 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.

FIGURE 5–5

Confidence bands for sample volatility.



Finally, $\hat{\sigma}$ can be used to estimate any quantile (an example is shown in Section 5.1.4). Since the normal distribution is fully characterized by two parameters only, the standard deviation contains all the information necessary to build measures of dispersion. Any σ-based quantile can be derived as

$$\hat{q}_{\sigma} = \alpha\hat{\sigma} \tag{5.20}$$

At the 95 percent confidence level, for instance, we simply multiply the estimated value of $\hat{\sigma}$ by 1.65 to find the 5 percent left-tail quantile. Of course, this method will be strictly valid if the underlying distribution is closely approximated by the normal. When the distribution is suspected to be strongly nonnormal, other methods, such as kernel estimation, also provide estimates of the quantile based on the full distribution.[1]

5.3.3 Estimation Error in Sample Quantiles

For arbitrary distributions, the cth quantile can be determined empirically from the historical distribution as $\hat{q}(c)$ (as shown in Section 5.1.2). There is, as before, some sampling error associated with the statistic. Kendall (1994) reports that the asymptotic standard error of $\hat{q}$ is

$$se(\hat{q}) = \sqrt{\frac{c(1-c)}{Tf(q)^2}} \tag{5.21}$$

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 5–6, where the expected quantile and 95 percent confidence bands are plotted for quantiles from the normal distribution.

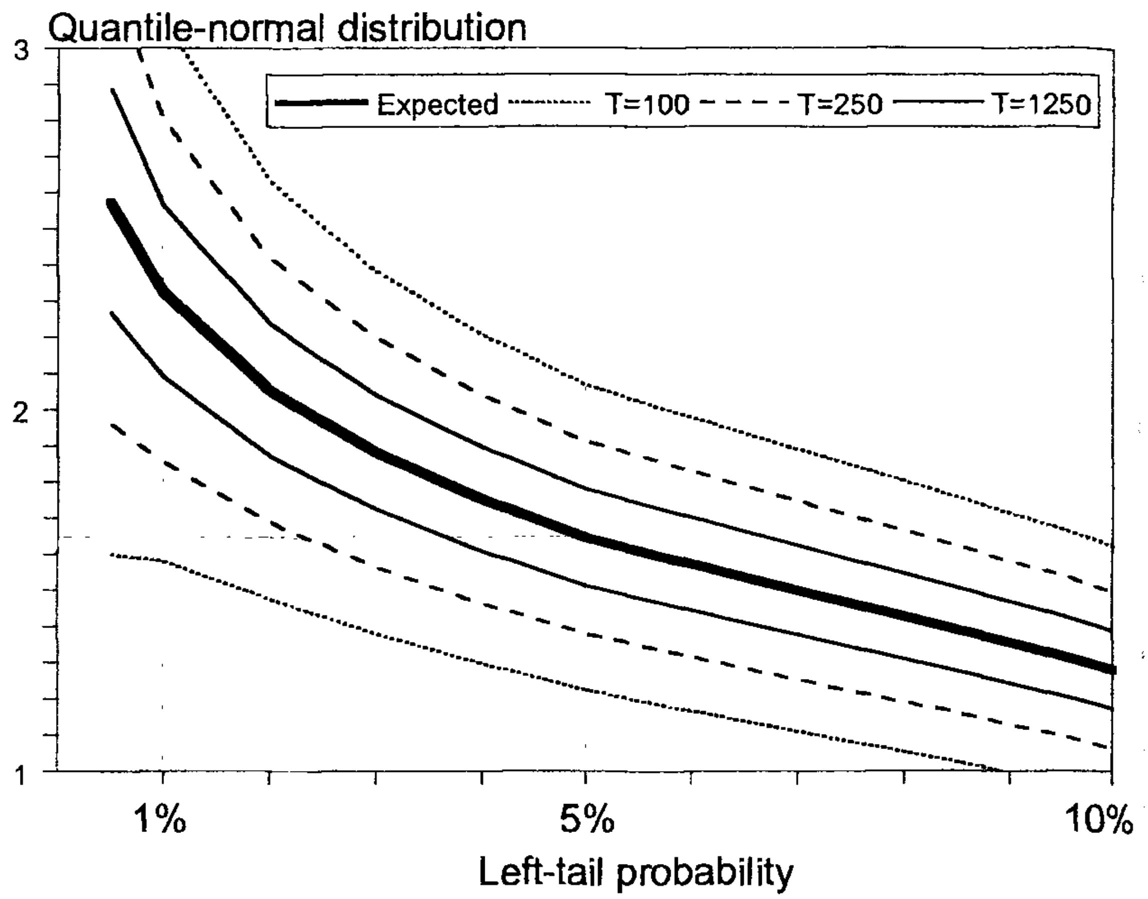
For the normal distribution, the 5 percent left-tailed interval is centered at 1.65. With $T = 100$, the confidence band is [1.24, 2.04], which is quite large. With 250 observations, which correspond to 1 year of trading days, the band is still [1.38, 1.91]. With $T = 1250$, or 5 years of data, the interval 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]. The interval of uncertainty is about

1. Kernel estimation smoothes the empirical distribution by a weighted sum of local distributions. For a further description of kernel estimation methods, see Scott (1992). Butler and Schachter (1998) apply this method to the estimation of VAR.

FIGURE 5–6

Confidence bands for sample quantiles.



twice that at the 5 percent interval. Thus sample quantiles are increasingly unreliable as one goes farther in the left tail.

As expected, 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.

5.3.4 Comparison of Methods

So far we have developed two approaches for measuring a distribution's VAR: (1) by directly reading the quantile 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, we may expect the σ-based approach to be more precise. Indeed, $\hat{\sigma}$ uses information about the whole distribution (in terms of all

TABLE 5-3

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

	VAR Confidence Level c	
	99%	95%
Exact quantile	2.33	1.65
Confidence band		
Sample $\hat{q}$	[1.85, 2.80]	[1.38, 1.91]
σ-Based, $\alpha\hat{\sigma}$	[2.24, 2.42]	[1.50, 1.78]

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 should still expect a performance improvement because the standard deviation uses all the sample information.

Table 5-3 compares 95 percent confidence bands for the two methods.[2] 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 much narrower than the previous interval.

A number of important conclusions can be derived from these numbers. First, there is substantial estimation error in the estimated quantiles, especially for high confidence levels, which are associated with rare events and hence difficult to verify. Second, parametric methods provide a substantial increase in precision, since the sample standard deviation contains far more information than sample quantiles.

Returning to the \$15.2 million VAR figure at the beginning of this chapter, we can now assess the precision of this number. Using the parametric approach based on a normal distribution, the standard error of this number is $se(\hat{q}_\sigma) = \alpha \times se(\hat{\sigma}) = 1.65 \times \$9.2 \text{ million } 1/(\sqrt{2 \times 254}) =$ \$0.67. Therefore, a two-standard-error confidence band around the VAR

2. For extensions to other distributions such as the Student, see Jorion (1996).

estimate is [$13.8 million, $16.6 million]. This narrow interval should provide reassurance that the VAR estimate is indeed meaningful.

5.4 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 is then 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 selection of the confidence level and horizon. On the one hand, if VAR is used simply as a benchmark risk measure, the choice is arbitrary and only needs to be consistent. 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.

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 risks. 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 one summary measure of risk.

As will be seen in Chapter 9, 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 method gives users much control over the measurement of risk, including 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 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 is then obtained, as described in Figure 7–2, as

$$\text{Incremental VAR} = VAR_{p+a} - VAR_p \tag{7.21}$$

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.

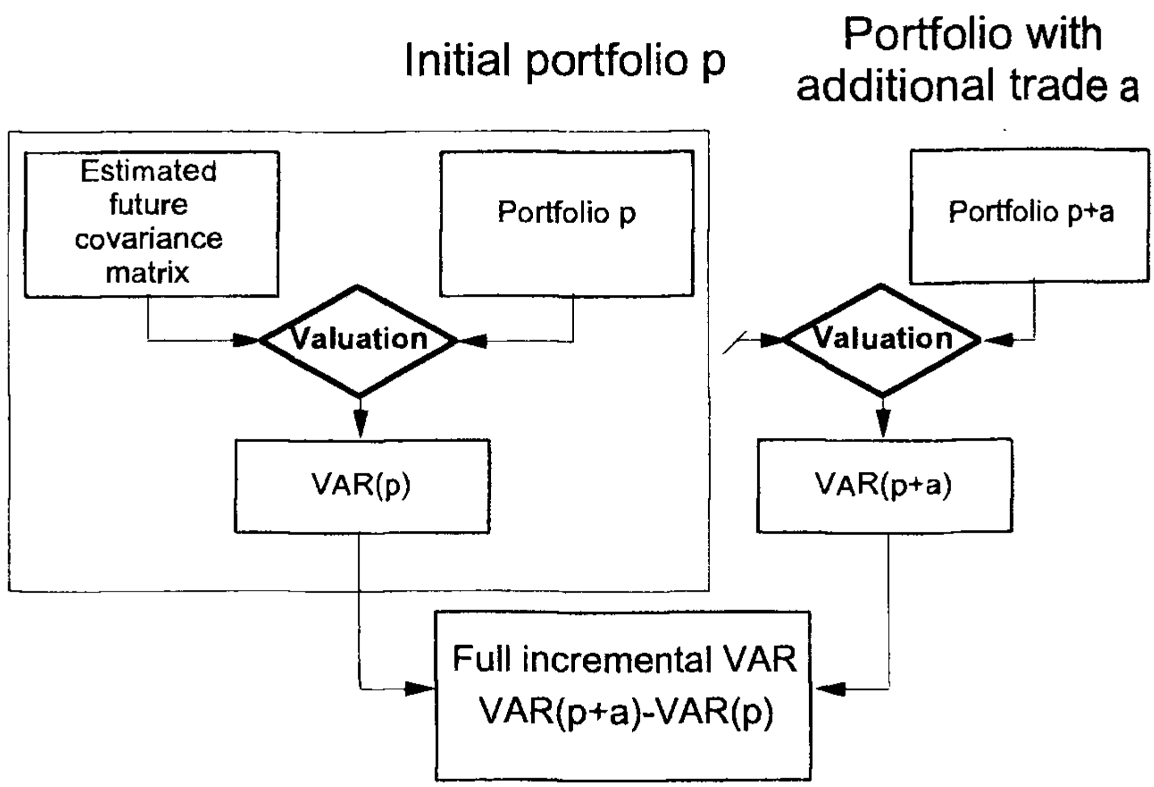
Incremental VAR

The change in VAR due 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-

FIGURE 7–2

The impact of a proposed trade with full revaluation.



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. Then a client comes with a proposed trade. Evaluating the effect of this trade on the bank's portfolio would again require 10 minutes using the incremental VAR approach. Most likely, this will be too long to wait to take action.

If we are willing to accept an approximation, however, we can take a shortcut.[1] Expanding VAR_{p+a} in series around the original point,

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

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

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

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.

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 assets. 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.

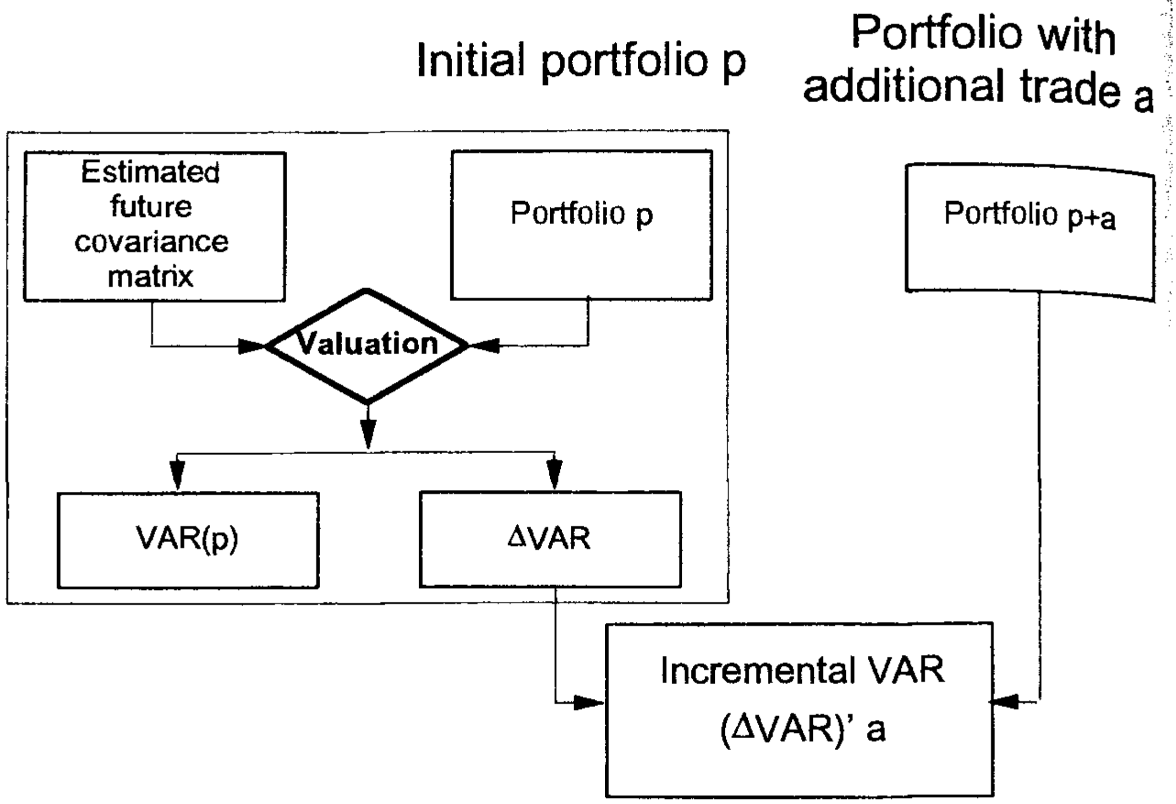
The incremental VAR method applies to the general case where a trade involves a set of new exposures on the risk factors. Consider instead the particular case where a new trade involves a position in one risk factor only (or asset). The portfolio value changes from the old value of W to the new value of $W_N = W + a$, where a is the amount invested in asset i. We can write the variance of the dollar returns on the new portfolio as

$$\sigma_N^2 W_N^2 = \sigma_p^2 W^2 + 2aW\sigma_{ip} + a^2\sigma_i^2 \qquad (7.24)$$

1. See also Garman (1996; 1997).

FIGURE 7–3

The impact of a proposed trade with marginal VAR.



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,

$$\frac{\partial \sigma_N^2 W_N^2}{\partial a} = 2W\sigma_{ip} + 2a\sigma_i^2 \tag{7.25}$$

which attains a zero value for

$$a^* = -W\frac{\sigma_{ip}}{\sigma_i^2} = -W\beta_i\frac{\sigma_p^2}{\sigma_i^2} \tag{7.26}$$

This is the variance-minimizing position, also known as *best hedge.*

Best hedge

Additional amount to invest in an asset so as to minimize the risk of the total portfolio.

Example (continued)

Going back to the previous two-currency example, we are now considering increasing the CAD position by US$10,000.

First, we use the marginal VAR method. We note that β can be obtained from a previous intermediate step as

$$\beta = \frac{\text{cov}(R_i,R_p)}{\sigma_p^2} = \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} / (\$0.156^2) = \begin{bmatrix} 0.205 \\ 0.590 \end{bmatrix}$$

The marginal VAR is now

$$\Delta\text{VAR} = \alpha \frac{\text{cov}(R_i,R_p)}{\sigma_p} = 1.65 \times \begin{bmatrix} \$0.0050 \\ \$0.0144 \end{bmatrix} / \$0.156 = \begin{bmatrix} 0.0528 \\ 0.1521 \end{bmatrix}$$

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 to the incremental VAR obtained from a full revaluation of the portfolio risk. We find

$$\sigma_{p+a}^2 = [\$2.01 \;\; \$1] \begin{bmatrix} 0.05^2 & 0 \\ 0 & 0.12^2 \end{bmatrix} \begin{bmatrix} \$2.01 \\ \$1 \end{bmatrix}$$

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:

$$\text{Component VAR} = (\Delta\text{VAR}_i) \times w_i\, W = \text{VAR}\beta_i\, w_i \qquad (7.27)$$

Thus the component VAR indicates how the portfolio VAR would change approximately if the component was deleted from the portfolio. We should note, however, that the quality of this linear approximation improves when the VAR components are small. Hence this decomposition is more useful with large portfolios, which tend to have many small positions.

We now show that these component VARs precisely add up to the total portfolio VAR. The sum is

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

because the term between parentheses is simply the beta of the portfolio with itself, which is unity.[2]

Thus we established that these *component VAR* measures add up to the total VAR. We have an additive measure of portfolio risk that reflects correlations. Components with a negative sign act as a hedge against the remainder of the portfolio. In contrast, components with a positive sign increase the risk of the portfolio.

Component VAR

A partition of the portfolio VAR that indicates how much the portfolio VAR would change approximately if the given component was deleted.

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

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

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

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

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,

2. This can be proved by expanding the portfolio variance into $\sigma_p^2 = w_1\text{cov}(R_1,R_p) + w_2\text{cov}(R_2,R_p) + \cdots = w_1(\beta_1\sigma_p^2) + w_2(\beta_2\sigma_p^2) + \cdots = \sigma_p^2(\Sigma_{i=1}^{N} w_i\beta_i)$ [thus the term between parentheses must be equal to one].

or by business unit. Such detail is invaluable for "drill down" exercises, which enable users to control their VAR.

Example (continued)

Continuing with the previous two-currency example, we find the component VAR for the portfolio using $\text{CVAR}_i = \Delta\text{VAR}_i x_i$,

$$\begin{bmatrix} \text{CVAR}_1 \\ \text{CVAR}_2 \end{bmatrix} = \begin{bmatrix} 0.0528 \times \$2 \text{ million} \\ 0.1521 \times \$1 \text{ million} \end{bmatrix} = \begin{bmatrix} \$105{,}630 \\ \$152{,}108 \end{bmatrix} = \text{VAR} \times \begin{bmatrix} 41.0\% \\ 59.0\% \end{bmatrix}$$

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.

Next, we can compute the change in the VAR if the EUR 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, $\text{VAR}_1 = \$165{,}000$. 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 as the VAR components are small relative to the total VAR.

7.2.4 Summary

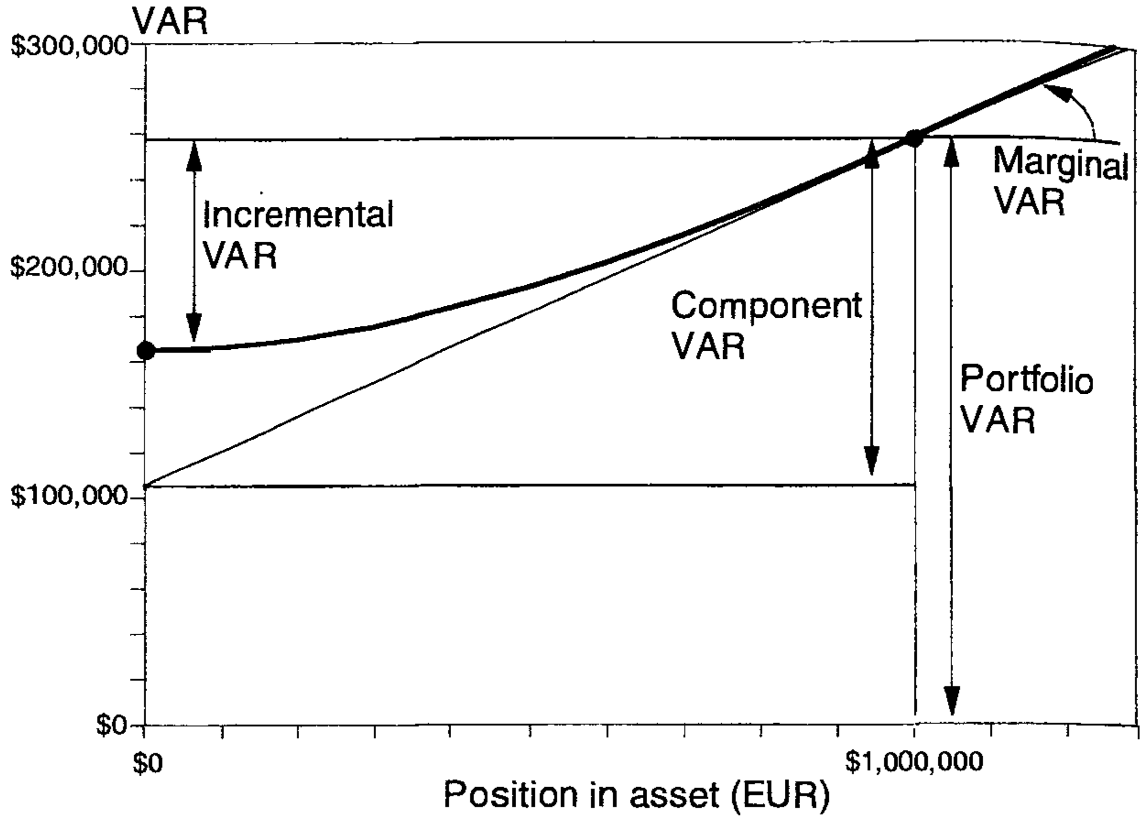
Figure 7–4 presents a graphic summary of our VAR tools for our two-currency portfolio. The graph plots the portfolio VAR as a function of the amount invested in this asset, the euro. At the current position of \$1 million, the portfolio VAR is \$257,738.

The marginal VAR is the change in VAR due to an addition of \$1 in EUR, or 0.1521; 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 due 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 to the incremental VAR. These component VAR measures add up to the total portfolio VAR, which gives a quick decomposition of the total risk.

FIGURE 7–4

VAR decomposition.



The graph also shows that the best hedge is a net zero position in the euro. Indeed, the VAR function attains a minimum when the position in the euro is zero.

The results are summarized in Table 7–1. This report gives not only the portfolio VAR but also a wealth of information for risk managers. For instance, the marginal VAR column can be used to determine how to reduce risk. Since the marginal VAR for the EUR is three times as large as that for the CAD, cutting the position in the EUR will be much more effective than cutting the CAD position by the same amount.

7.3 EXAMPLES

This section provides a number of applications of VAR measures. The first example illustrates a risk report for a global equity portfolio. The second shows how VAR could have been used to dissect the Barings portfolio.

TABLE 7–1

VAR Decomposition for Sample Portfolio

Currency	Current Position, x_i or $w_i W$	Individual VAR, $VAR_i = \alpha\sigma_i w_i W$	Marginal VAR, $\Delta VAR_i = VAR\ \beta_i / W$	Component VAR, $CVAR_i = \Delta VAR_i x_i$	Percent Contribution, $CVAR_i / VAR$
CAD	$2 million	$165,000	0.0528	$105,630	41.0%
EUR	$1 million	$198,000	0.1521	$152,108	59.0%
Total	$3 million				
Undiversified VAR		$363,000			
Diversified VAR				$257,738	100.0%

7.3.1 A Global Portfolio Equity Report

To further illustrate the use of our VAR tools, Table 7–2 displays a risk management report for a global equity portfolio. Here, risk is measured in relative terms, i.e., relative to the benchmark index. The current portfolio has an annualized tracking error, which is also the standard deviation of the difference 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 σ_p.

Positions are reported as deviations in percent from the benchmark in the second column. Since the weights of the index and of the current portfolio must sum to one, the deviations must sum to zero.

The next columns report the individual risk, marginal risk, and the 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, since the positions in these markets, displayed in the first column, are not the largest. 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 contribution of Japan and Brazil is high because of their high volatility and correlations with the portfolio.

3. Hot Spots is a trademark of Goldman Sachs.

TABLE 7–2

Global Equity Portfolio Report

Country	Current Position (%) w_i	Individual Risk $w_i\sigma_i$	Marginal Risk β_i	Percentage 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		
volatility:						
Undiversified		6.91%				
Diversified	1.82%					

Source: Adapted from Litterman (1996).

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 percent 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 can be quite useful, however, because analysts can now balance their return forecasts against risk explicitly.

7.3.2 Barings: An Example in Risks

Barings' collapse provides an interesting application of the VAR methodology. Leeson was reported to be long about $7.7 billion worth of Japanese stock index (Nikkei) futures and short $16 billion worth of Japanese Government Bond (JGB) futures. Unfortunately, official reports to Barings showed "nil" risk because the positions were fraudulent.

If a proper VAR system had been in place, the parent company could have answered the following questions. What was Leeson's actual VAR? Which component contributed most to VAR? Were the positions hedging each other or adding to the risk?

The top panel of Table 7–3 displays monthly volatility measures and correlations for positions in the 10-year zero JGB and the Nikkei index. The correlation between Japanese stocks and bonds is negative, indicating that increases in stock prices are associated with decreases in bond prices or increases in interest rates. The next column displays positions that are reported in millions of dollar equivalents.

To compute the VAR, we first construct the covariance matrix Σ from the correlations. Next, we compute the vector Σx, which is in the first column of the bottom panel. For instance, the -2.82 entry is found from $\sigma_1^2 x_1 + \sigma_{12} x_2 = 0.000139 \times (-\$16{,}000) + (-0.000078) \times \$7700 = -2.82$. The next column reports $x_1(\Sigma x)_1$ and $x_2(\Sigma x)_2$, which sum to the total portfolio variance of 256,193.8, for a portfolio volatility of $\sqrt{256{,}194} = \$506$ million. At the 95 percent confidence level, Barings' VAR was $\$1.65 \times \$506$, or $835 million.

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 with the VAR reported here.

TABLE 7–3

Barings' Risks

	Risk (%) σ	Correlation Matrix		Covariance Matrix Σ		Positions (\$ millions) x	Individual VAR $\alpha\sigma x$
10-yr 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	\$1,051.39

	Total VAR		Marginal VAR			
Asset i	$(\Sigma x)_i$	$x'_i(\Sigma x)_i$	β_i $(\Sigma x)_i/\sigma_p^2$	For \$1 million β_i VAR	Component VAR $\beta_i x_i$ VAR	Percent Contribution
10-yr JGB	−2.82	45,138.8	−0.0000110	(\$0.00920)	\$147.15	17.6%
Nikkei	27.41	211,055.1	0.0001070	\$0.08935	\$688.01	82.4%
Total		256,193.8			\$835.16	100.0%
Risk = σ_p		506.16				
VAR = $\alpha\sigma_p$		\$835.16				

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 due to 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 due to the total bond position is \$147.15 million; that due to the stock position is \$688.01 million. By construction, these two numbers add up to the total VAR of \$835.16 million. The percent contributions are reported in the last column. This analysis reveals that most of the loss was due to the Nikkei exposure and that the bond position, instead of hedging, made things even worse.

7.4 SIMPLIFYING THE COVARIANCE MATRIX

7.4.1 Why Simplifications?

So far we have shown that correlations are essential driving forces behind portfolio risk. When the number of assets is large, however, the measurement of the covariance matrix becomes increasingly difficult. With 10 assets, for instance, we need to estimate $10 \times 11/2 = 55$ different variance and covariance terms. With 100 assets, this number climbs to 5050. The number of correlations increases geometrically with the number of

4. *Financial Times* (March 1, 1995).

assets. For large portfolios, this causes real problems: (1) The portfolio VAR may not be positive, and (2) correlations may be estimated imprecisely. This section examines the extent to which such problems can affect VAR measures and proposes some solutions.

In practice, the industry has developed a number of approximations to the covariance matrix. Securities are mapped routinely over general risk factors. These *mapping* procedures replace the exposure profile of a security by that of appropriately chosen indices. The latter are called *general market risks*. In contrast, the remaining risks are called *specific risks*.

Mapping cuts down the computational requirements when there is a large number of positions. In some situations, also, we may not have a complete history of securities data, in which case mapping provides a useful replacement for the security.

7.4.2 Zero VAR Measures

The VAR measure derives from the portfolio variance, which is computed as

$$\sigma_p^2 = w'\Sigma w \tag{7.31}$$

The question is, Is this product guaranteed to be always positive?

Unfortunately, not always.[5] For this to be the case, we need the matrix Σ to be *positive definite*. This can be verified using the singular value decomposition described in Appendix 7B.

A number of conditions must be satisfied for positive-definiteness: The number of historical observations T must be greater than the number of assets N, and the series cannot be linearly correlated. The first condition states that if a portfolio consists of 100 assets, there must be at least 100 historical observations to ensure that whatever portfolio is selected, the portfolio variance will be positive. The second condition rules out situations where an asset is exactly equivalent to a linear combination of other assets.

An example of a non-positive-definite matrix is obtained 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 cur-

5. Abstracting from the obvious case where all elements of w are zero.

rencies fixed to each other). In addition, positions must have been matched precisely with assets so as to yield zero risk. This is most likely to occur if the weights have 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.

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.

√ 7.4.3 Diagonal Model

A related problem is that as the number of assets increases, it is more likely that some correlations will be measured with error. Some models can help simplifying this process by providing a simpler structure for the covariance matrix. One such model is the *diagonal model,* originally proposed by Sharpe in the context of stock portfolios.[6]

The assumption is that the common movement in all assets is due to one common factor only, the market. Formally, the model is

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i \qquad E[\epsilon_i] = 0 \qquad E[\epsilon_i R_m] = 0$$
$$E[\epsilon_i \epsilon_j] = 0 \qquad E[\epsilon_i^2] = \sigma^2_{\epsilon,i} \quad (7.32)$$

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 can be decomposed as

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma^2_{\epsilon,i} \quad (7.33)$$

The covariance between two assets is

$$\sigma_{i,j} = \beta_i \beta_j \sigma_m^2 \quad (7.34)$$

which is solely due to the common factor. The full matrix is

$$\Sigma = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} [\beta_1 \cdots \beta_N] \sigma_m^2 + \begin{bmatrix} \sigma^2_{\epsilon,1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma^2_{\epsilon,N} \end{bmatrix}$$

6. Note that this model is sometimes referred to as the CAPM, which is not correct. The diagonal model is simply a simplification of the covariance matrix and says nothing about expected returns, whose description is the essence of the CAPM.

Written in matrix notation, the covariance matrix is

$$\Sigma = \beta\beta'\sigma_m^2 + D_\epsilon \tag{7.35}$$

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 is reduced from 5050 to 201, a considerable improvement.

Furthermore, the variance of large, well-diversified portfolios simplifies even further, reflecting only exposure to the common factor. The variance of the portfolio is

$$V(R_p) = V(w'R) = w'\Sigma w = (w'\beta\beta'w)\sigma_m^2 + w'D_\epsilon w \tag{7.36}$$

The second term consists of $\Sigma_{i=1}^{N} w_i^2 \sigma_{\epsilon,i}^2$. But this term becomes very small as the number of securities in the portfolio increases. For instance, if all the residual variances are identical and have equal weights, this second term is $[\Sigma_{i=1}^{N}(1/N)^2]\sigma_\epsilon^2$, which converges to 0 as N increases. Therefore, the variance of the portfolio converges to

$$V(R_p) \rightarrow (w'\beta\beta'w)\sigma_m^2 = (\beta_p\sigma_m)^2 \tag{7.37}$$

which depends on one factor only. Thus, in large portfolios, specific risk becomes unimportant for the purpose of measuring VAR.

As an example, consider three stocks, General Motors (GM), Ford, and Hewlett-Packard (HWP). The top panel in Table 7–4 displays the full covariance matrix for monthly data. 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; HWP has the highest systematic risk. The market variance is $V(R_m) = 11.90$. 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 $0.806^2 \times 11.90 + 64.44 = 7.73 + 64.44 = 72.17$. The covariance between GM and Ford is $\beta_1\beta_2 V(R_m)$, which is $0.806 \times 1.183 \times 11.90 = 11.35$.

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

TABLE 7–4

The Diagonal Model

	Covariance			Correlations		
	GM	Ford	HWP	GM	Ford	HWP
Full matrix						
GM	72.17			1		
Ford	43.92	66.12		0.636	1	
HWP	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	
HWP	17.87	26.23	90.41	0.221	0.339	1

model, the correlation is driven by exposure to the market, and is 0.164, which is lower than the true correlation. This is so because both stocks have low betas, which is the only source of common variation. Whether this model produces acceptable approximations depends on the purpose at hand; we compute actual VAR numbers in Chapter 11. But there is no question that the diagonal model provides a considerable simplification.

7.4.4 Factor Models

If a one-factor model is not sufficient, better precision can be obtained with multiple factor models. Equation (7.32) can be generalized to K factors

$$R_i = \alpha_i + \beta_{i1} y_1 + \cdots + \beta_{iK} y_K + \epsilon_i \tag{7.38}$$

where $R_1, \ldots, R_N$ are the N asset returns and $y_1, \ldots, y_K$ 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 the transportation industry, that would pick up the higher correlation between GM and Ford. With multiple factors, the covariance matrix acquires a richer structure

$$\Sigma = \beta_1\beta_1'\sigma_1^2 + \cdots + \beta_K\beta_K'\sigma_K^2 + D_\epsilon \tag{7.39}$$

The total number of parameters is $(N \times K + K + N)$, which may still be considerably less 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.

Factor models are also important because they can help us 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?

To illustrate, consider the U.S. Treasury bond market. Table 7–5 presents monthly VARs for 11 zero-coupon bonds as well as correlations for maturities going from 1 to 30 years. Under the RiskMetrics convention, VAR corresponds to 1.65 standard deviations. With strictly parallel moves in the term structure, VAR should increase linearly with maturity. In fact, this is not the case. Longer maturities display slightly less VAR than under a linear relationship. The 30-year zero, for instance, has a VAR of 11.12 instead of the value of 14.09 extrapolated from the 1-year maturity ($0.470 \times 30/1$).

Particularly interesting are the very high correlations, confirming the presence of one major factor behind bond returns. Correlations are 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?

To answer this question, we can turn to the principal-components method. Intuitively, *principal components* attempts to find a series of independent linear combinations of the original variables that provides the best explanation of diagonal terms of the matrix. The methodology is summarized in Appendix 7B.

Another statistical method is *factor analysis*. The latter uses maximum-likelihood techniques to estimate the factor loadings under the restriction that the residual matrix is diagonal and assuming that returns are normally distributed. Factor analysis differs from principal component in that it focuses on the off-diagonal elements of the correlation matrix.

TABLE 7–5

Risk and Correlations for U.S. Bonds (Monthly VAR at 95 Percent Level)

Term (year)	VAR (%)	1Y	2Y	3Y	4Y	5Y	7Y	9Y	10Y	15Y	20Y	30Y
1	0.470	1										
2	0.967	0.897	1									
3	1.484	0.866	0.991	1								
4	1.971	0.866	0.976	0.994	1							
5	2.426	0.855	0.966	0.968	0.996	1						
7	3.192	0.825	0.936	0.965	0.982	0.990	1					
9	3.913	0.796	0.909	0.942	0.064	0.975	0.996	1				
10	4.250	0.788	0.903	0.937	0.959	0.971	0.994	0.999	1			
15	6.234	0.740	0.853	0.691	0.915	0.930	0.961	0.976	0.981	1		
20	8.146	0.679	0.791	0.832	0.660	0.878	0.919	0.942	0.951	0.991	1	
30	11.119	0.644	0.761	0.801	0.831	0.853	0.902	0.931	0.943	0.975	0.986	1

Which method is best depends on the purpose at hand. Wilson (1994) argues that principal components should be used for applications that rely on accurate modeling of volatility, such as simple options. On the other hand, applications for which correlations are critical, such as "diff" swaps, would benefit from factor analysis.

Table 7–6 provides an answer. It shows the first three components for the correlation matrix of U.S. bond returns, based on principal-component analysis. The most striking feature of the table is that the first factor provides an excellent fit to movements of the term structure. The average explanatory power is very high, at 91.9 percent. Since it affects all maturities about equally, this common factor can be defined a yield *level* variable. This explains why duration is a good measure of interest rate risk.

The second factor explains an additional 6.0 percent in movements. Since 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

TABLE 7–6

Principal Components of Correlation Matrix: U.S. Bonds

Maturity (year)	*Percentage of Variance Explained by* Factor 1, Level	Factor 2, Slope	Factor 3	Total Variance Explained
1	72.2	17.9	9.8	99.8
2	89.7	7.8	0.5	98.0
3	94.3	4.5	0.7	99.5
4	96.5	2.2	1.0	99.7
5	97.7	1.1	0.9	99.7
7	98.9	0.0	0.4	99.3
9	98.2	0.7	0.2	99.1
10	98.1	1.2	0.1	99.4
15	94.1	5.3	0.2	99.6
20	87.2	11.0	0.9	99.1
30	83.6	14.5	0.9	99.0
Average	91.9	6.0	1.4	99.3

1-year rates, perhaps because of different characteristics of money market instruments. Together, the three factors explain an impressive 99.3 percent of all return variation.

This decomposition shows that the risk of a bond portfolio can be usefully summarized by its exposure to two factors only. For instance, the portfolio can be structured so that the net exposure to the two factors is very small. This will improve considerably on duration hedging yet require no forecast of future twists in the yield curve. In other words, we need only two primitive risk factors to represent movements in the yield curve.

7.4.5 Comparison of Methods

The purpose of these various methods is to simplify the computation of the portfolio risk. With hundreds of securities in the portfolio, it may not be feasible to consider each one as an individual risk factor. The question is whether these shortcuts materially affect the VAR measure.

To illustrate, Table 7–7 presents VAR calculations for three portfolios.[7] 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 5 stocks and short the others. VAR is measured with a 1-month horizon at the 95 percent level of confidence using historical data from 1990 to 1999.

To summarize, five methods are examined:

- *Index mapping* replaces each stock by a like position in the index m:

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

- *Beta mapping* also considers the net beta of the portfolio:

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

- *Diagonal model* considers both the beta and specific risk:

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

7. The diversified portfolio consists of positions in Ford, Hewlett-Packard, General Electric, Procter & Gamble, AT&T, Boeing, General Motors, Disney, Microsoft, and American Express. These are spread among 6 of the 10 industrial sectors in the market. The long-short portfolio is long the first five and short the others. The market index is taken as the S&P 500. This example is similar to that of Beder et al. (1998).

TABLE 7–7

Comparison of VAR Methods

Position	Portfolio: Diversified $1,000,000	Hi-Tech $1,000,000	Long-Short $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	$78,994	$118,955	$32,598

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

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

- *Individual mapping* uses the full covariance matrix of individual stocks:

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

This method provides an exact VAR measure over this sample period.

More complex models are certainly possible. For instance, one could model a marketwide effect, then industry effects, and finally assume that remaining terms are uncorrelated.

The table shows that the quality of the approximation depends on the structure of the portfolio. 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. The diagonal model gets close to the true value, as before.

Finally, the third portfolio shows the dangers of simple mapping methods. The index mapping model, given a zero net investment, 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.

Otherwise, a shortcut is sometimes used for portfolios with long and short positions. This consists of grouping the long and short positions separately and computing their VAR using beta mapping. Define these as VAR_L and VAR_S. The portfolio VAR is computed as

$$VAR = |VAR_L - VAR_S| + k \times \min(VAR_L, VAR_S) \qquad (7.40)$$

where k is an "offsetting factor." When $k = 0$, for instance, there is full offset between the two, which gives $VAR = VAR_L - VAR_S = \$298$. When $k = 2$, there is no offset at all between specific risks, which gives a very conservative measure, $VAR = VAR_L + VAR_S = \$70{,}086$. We have partial offset with $k = 1$, which gives $VAR = \max(VAR_L, VAR_S)$, or here, $VAR = VAR_L = \$35{,}192$, which is not far from the actual value. This ad hoc method is sometimes useful to deal with hedge portfolios.

7.5 CONCLUSIONS

Much of portfolio analysis is based on the fact that fixed portfolios of normally distributed variables are themselves normal. This leads to an analytical approach based on the covariance matrix that is particularly convenient for VAR calculations.

Indeed, closed-form solutions allow us to compute the marginal effect of changing a position and to decompose the current risk into additive components. Armed with these tools, users can better understand and manage their risks.

The disadvantage of these methods is that the covariance matrix quickly increases with the number of securities. This can cause computational problems. As a result, it has become common to model securities risk by "mapping" them over a smaller set of general risk factors.

Mapping, however, forfeits resolution. In most cases, the lack of detail is harmless. We have seen, for example, that for long portfolios, simple mapping methods generally produce acceptable results. Serious shortcomings arise, however, when the portfolio is hedged or uses long and short positions. In such a case, there is a potential for the manipulation of VAR, since simple methods are clearly unable to measure risk properly. This is an example of "gaming the VAR system," which is discussed in a later chapter. In summary, the application of these techniques should be adapted carefully to the portfolio at hand.

APPENDIX 7A: MATRIX MULTIPLICATION

This appendix reviews the algebra for matrix multiplication. Suppose 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).

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$C = AB = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

To multiply the matrix A by B, we compute each element by taking each row of A and multiplying by the desired column of B. For instance, element c_{ij} would be obtained by multiplying each element of the ith row of A individually by each element of the jth column of B and summing over all of 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

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{21} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{22} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

APPENDIX 7B (ADVANCED): 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 $z = w'R$ using $V^*(z) = w'\Sigma^* w$. The process consists of replacing the original variables R by another set y suitably selected.

The *first* principal component is the linear combination

$$y_1 = \beta_{11}R_1 + \cdots + \beta_{N1}R_N = \beta_1'R \tag{7.41}$$

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, $\sigma^2(y_1) = \beta_1'\Sigma\beta_1$, shows that the vector β_1 must satisfy $\Sigma\beta_1 = \lambda_1\beta_1$. Here, $\sigma^2(y_1) = \lambda_1$ is the largest *eigenvalue* of the matrix Σ, and β_1 its associated *eigenvector.*

The *second* principal component is the one that has greatest variance subject to the same normalization constraint $\beta_2'\beta_2 = 1$ and to the fact that is 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 y 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

$$\Sigma = PDP' = [\beta_1 \ldots \beta_N] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_N' \end{bmatrix} \tag{7.42}$$

where P is an orthogonal matrix, i.e., such that its inverse is also its transpose, $PP' = I$ (or $P^{-1} = P'$) 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 $y = P'R$. Alternatively, if we are given the set of y, we can recover R as $R = Py$. In other words,

$$R_i = \beta_{i1}y_1 + \cdots + \beta_{iN}y_N \tag{7.43}$$

To each y_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 properly defined. The problem is that for some portfolios, the resulting VAR could be negative![8]

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:

$$R_i \approx \beta_{i1}y_1 + \cdots + \beta_{iK}y_K \tag{7.44}$$

Based on this, we approximate the matrix by

$$\Sigma^* = [\beta_1 \ldots \beta_K] \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & \lambda_K \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_K' \end{bmatrix}$$

$$= \beta_1\beta_1'\lambda_I + \cdots + \beta_K\beta_K'\lambda_K \tag{7.45}$$

which is very close to Equation (7.39) except for the residual terms on the diagonal. Note that this matrix Σ^* is surely not invertible, since 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 y's called *principal components* (PCs).

---

8. It is possible to transform the matrix in a systematic fashion so that it avoids being non-positive definite. For a review of methods, see Rebonato and Jäckel (2000).

Given a portfolio $z = w'R$, the portfolio can be mapped into its exposures on the principal components:

$$\begin{aligned} z = \Sigma w_i R_i &\approx w_1(\beta_{11}y_1 + \cdots + \beta_{1K}y_K) + \cdots + w_N(\beta_{N1}y_1 + \cdots + \beta_{NK}y_K) \\ &= (w_1\beta_{11} + \cdots + w_N\beta_{N1})y_1 + \cdots + (w_1\beta_{1K} + \cdots + w_N\beta_{NK})y_K \\ &= \delta_1 y_1 + \cdots + \delta_K y_K \end{aligned}$$

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 (7.45):

$$\begin{aligned} \sigma^2(z) = w'\Sigma^* w &= w'\beta_1\beta_1' w\lambda_1 + \cdots + w'\beta_K\beta_K' w\lambda_K \\ &= (w'\beta_1)^2\lambda_1 + \cdots + (w'\beta_K)^2\lambda_K \\ &= \delta_1^2\sigma^2(y_1) + \cdots + \delta_K^2\sigma^2(y_K) \end{aligned} \tag{7.46}$$

which is remarkably simple. The variance of the portfolio z is given by sum of the squared exposures δ times the variance of each PC.

Instead of having to deal with all the variances and covariances of R, we simply use K independent terms. For instance, as in the example of a bond market, we can replace a covariance matrix of dimension 11 times 11 with 66 terms by 3 terms in all.

CHAPTER 9

VAR Methods

In practice, this works, but how about in theory?

Attributed to a French mathematician

Value at risk (VAR) has become an essential component in the toolkit of risk managers because it provides a quantitative measure of downside risk. In practice, the objective should be to provide a reasonably accurate estimate of risk at a reasonable cost. This involves choosing 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.

Approaches to VAR basically 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. The delta-normal method uses linear, or delta, derivatives and assumes normal distributions. Because the delta-normal approach is easy to implement, a variant, called the "Greeks," is sometimes used. This method consists of analytical approximations to first- and second-order derivatives and is most appropriate for portfolios with limited sources of risk. The second group uses full valuation. *Full-valuation methods* measure risk by fully repricing the portfolio over a range of scenarios. The pros and cons of local versus full valuation are discussed in Section 9.1. 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 best example of local valuation is the delta-normal method, which is explained in Section 9.2. Full valuation is implemented in the historical simulation method and the Monte Carlo simulation method, which are discussed in Sections 9.3 and 9.4.

This classification reflects a fundamental tradeoff between speed and accuracy. Speed is important for large portfolios exposed to many risk factors, which 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.

An in-depth analysis of the delta-normal and simulation VAR methods is presented in following chapters, as well as a related method, stress testing. Section 9.5 presents some empirical comparisons. Finally, Section 9.6 summarizes the pros and cons of each method.

9.1 LOCAL VERSUS FULL VALUATION

9.1.1 Delta-Normal Valuation

Local-valuation methods usually rely on the normality assumption for the driving risk factors. This assumption is particularly convenient because of the invariance property of normal variables: Portfolios of normal variables are themselves normally distributed.

We initially focus on *delta valuation,* which considers only the first derivatives. To illustrate the approaches, take an instrument whose value depends on a single underlying risk factor S. The first step consists of valuing the portfolio at the initial point

$$V_0 = V(S_0) \tag{9.1}$$

along with analytical or numerical derivatives. Define Δ_0 as the first partial derivative, or the portfolio sensitivity to changes in prices, evaluated at the current position V_0. This would be called *modified duration* for a fixed-income portfolio or delta for a derivative. 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 underlying asset. The portfolio Δ simply can be computed as the sum of individual deltas.

The potential loss in value dV is then computed as

$$dV = \frac{\partial V}{\partial S}\Big|_0 dS = \Delta_0 \times dS \tag{9.2}$$

which involves the potential change in prices dS. 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:

$$\text{VAR} = |\Delta_0| \times \text{VAR}_S = |\Delta_0| \times (\alpha\sigma S_0) \tag{9.3}$$

where α is the standard normal deviate corresponding to the specified confidence level, e.g., 1.645 for a 95 percent confidence level. Here, we take $\sigma(dS/S)$ as the standard deviation of *rates* of changes in the price. The assumption is that rates of changes are normally distributed.

Because VAR is obtained as a closed-form solution, this method is called *analytical*. Note that VAR was measured by computing the portfolio value only once, at the current position V_0.

For a fixed-income portfolio, the risk factor is the yield y, and the price-yield relationship is

$$dV = -D^*V\,dy \tag{9.4}$$

where D^* is the *modified duration*. In this case, the portfolio VAR is

$$\text{VAR} = (D^*V) \times (\alpha\sigma) \tag{9.5}$$

where $\sigma(dy)$ is now the volatility of changes in the *level* of yield. The assumption is that changes in yields are normally distributed, although this is ultimately an empirical issue.

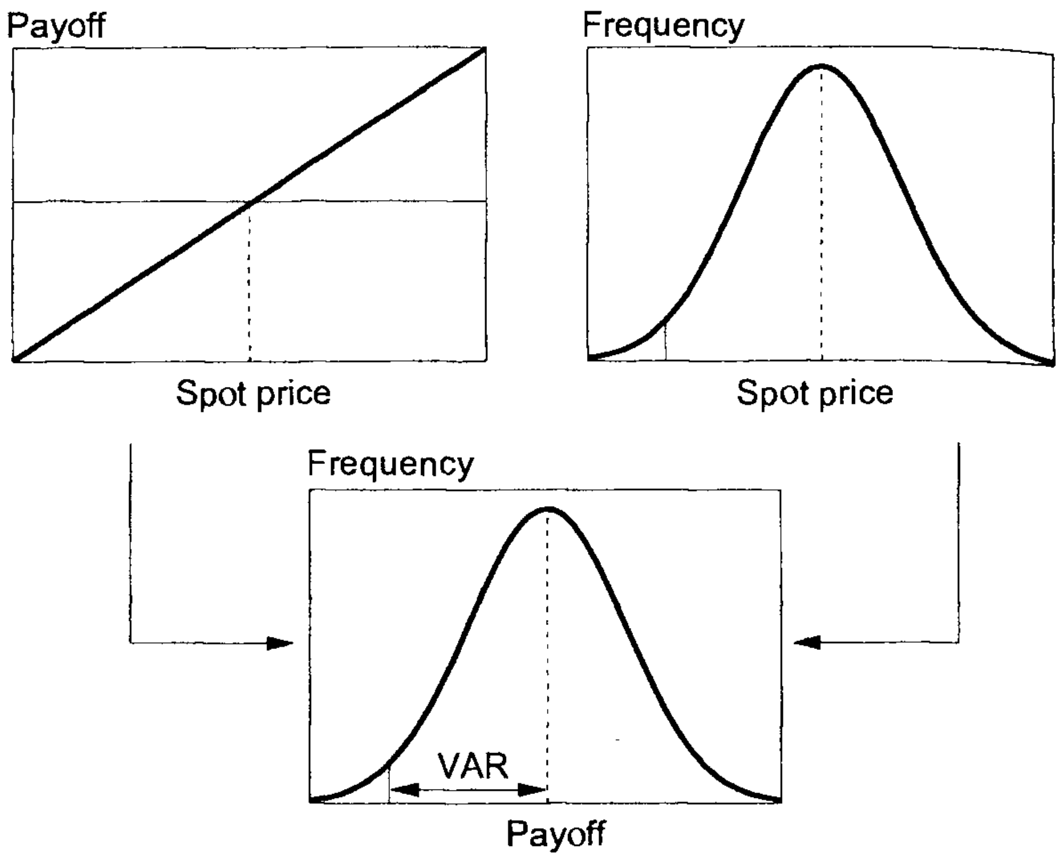
This method is illustrated in Figure 9–1, where the profit payoff is a linear function of the underlying spot price and is displayed at the upper left side; the price itself is normally distributed, as shown in the right panel. As a result, the profit itself is normally distributed, as shown at the bottom of the figure. The VAR for the profit can be found from the exposure and the VAR for the underlying price. There is a one-to-one mapping between the two VAR measures.

How good is this approximation? It depends on the "optionality" of the portfolio as well as the horizon. Consider, for instance, a simple case of a long position in a call option. In this case, we can easily describe the distribution of option values. 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.

This is illustrated in Figure 9–2, 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, due to the upside potential, whereas the downside is limited to the option premium. This shift is due to the nonlinear payoff on the option.

FIGURE 9-1

Distribution with linear exposures.



Here, the cth quantile for V is simply the function evaluated at the cth quantile of S. For the long-call option, the worst loss for V at a given confidence level will be achieved at $S^* = S_0 - \alpha\sigma S_0$, and

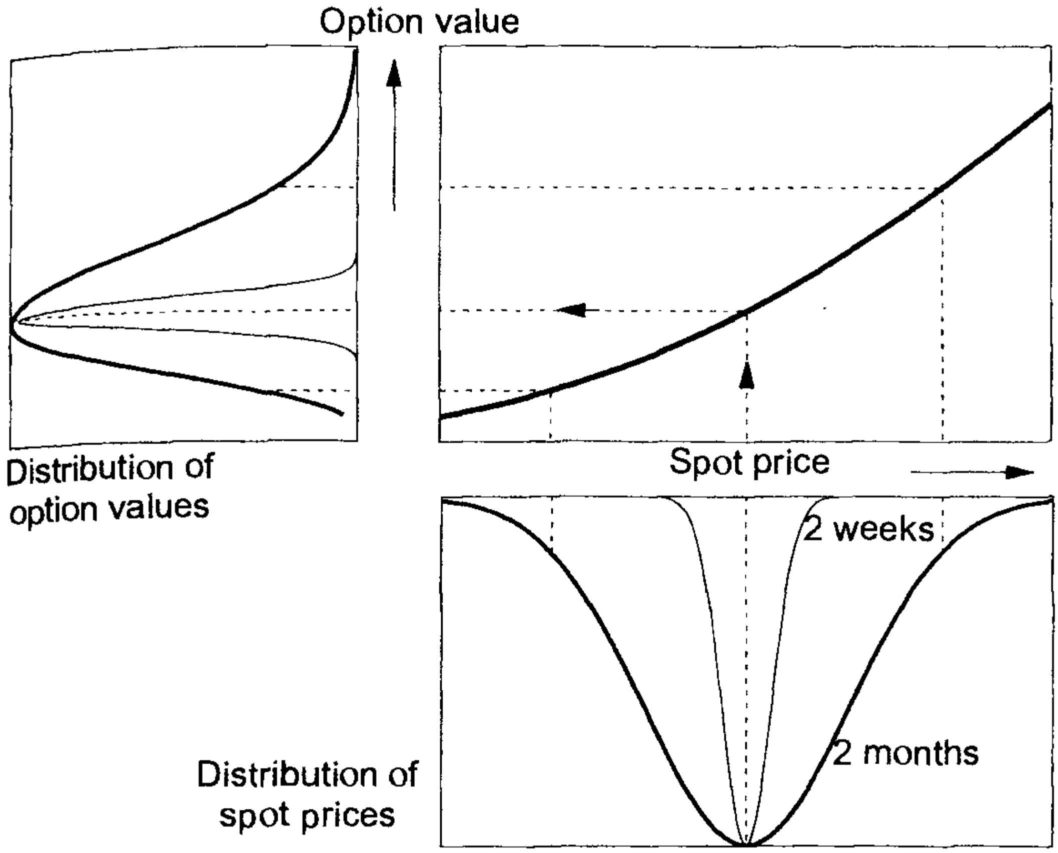
$$\text{VAR} = V(S_0) - V(S_0 - \alpha\sigma S_0) \tag{9.6}$$

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

FIGURE 9–2

Transformation of distributions.



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.

9.1.2 Full Valuation

In some situations, the delta-normal approach is totally inadequate. This is the case, for instance, when the worst loss may not be obtained for extreme realizations of the underlying spot rate. Also, options that are near expiration and at-the-money have unstable deltas, which translate into asymmetrical payoff distributions.

An example of this problem 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:

$$dV = V(S_1) - V(S_0) \tag{9.7}$$

The new values S_1 can be generated by simulation methods. The *Monte Carlo simulation approach* relies on prespecified distributions. For instance, the realizations can be drawn from a normal distribution,

$$dS/S \approx N(0, \sigma^2) \tag{9.8}$$

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

For each of these draws, the portfolio is priced on the target date using a full-valuation method. This method is potentially the most accurate because it accounts for nonlinearities, income payments, and even time-decay effects that are usually ignored in the delta-normal approach. VAR is then calculated from the percentiles of the full distribution of payoffs. 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.

To illustrate the result of nonlinear exposures, Figure 9–3 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.

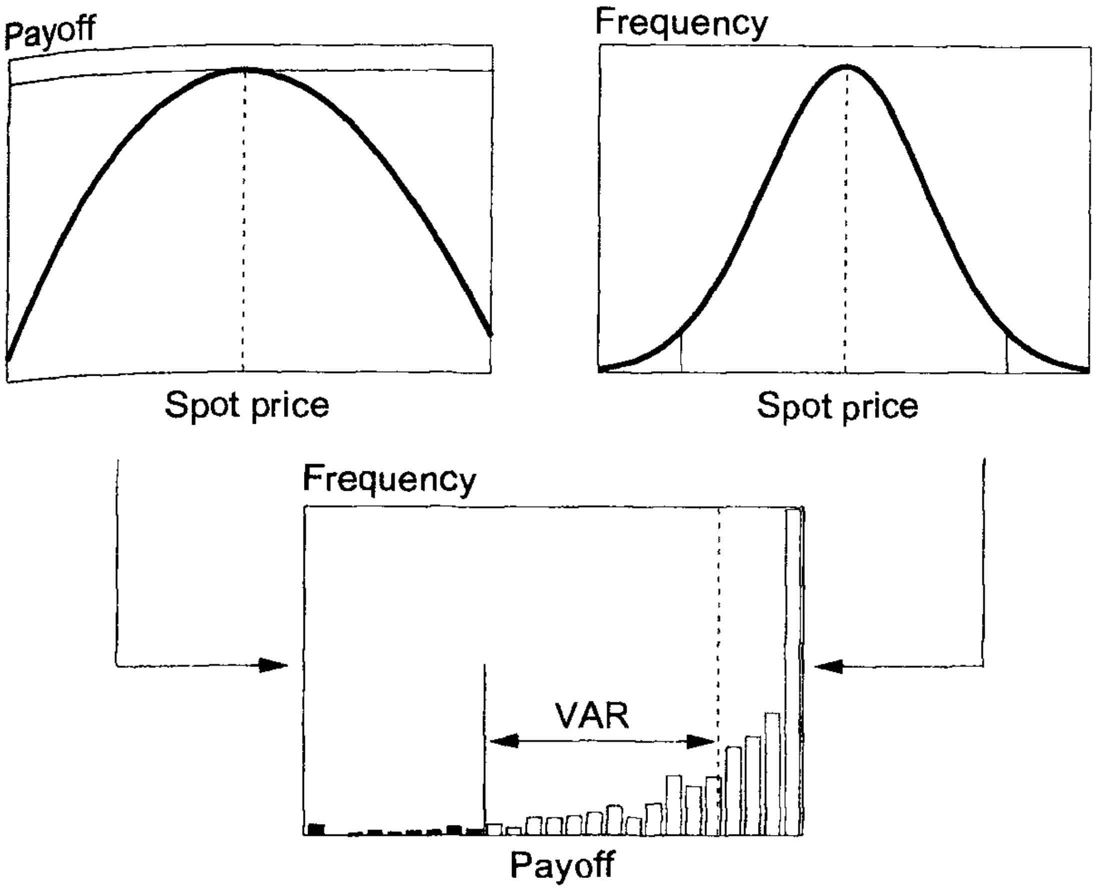
The problem is that these simulation methods require substantial computing time when applied to large portfolios. As a result, methods have been developed to speed up the computations.

One example is the *grid Monte Carlo approach,* which starts by an exact valuation of the portfolio over a limited number of grid points.[1] For each simulation, the portfolio value is then 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.

1. Picoult (1997) describes this method in more detail.

FIGURE 9–3

Distribution with nonlinear exposures.



9.1.3 Delta-Gamma Approximations (the "Greeks")

It may be possible to extend the analytical tractability of the delta-normal method with higher-order terms. We can improve the quality of the linear approximation by adding terms in the Taylor expansion of the valuation function:

$$dV = \frac{\partial V}{\partial S}\,dS + \frac{1}{2}\frac{\partial^2 V}{\partial S_2}\,dS^2 + \frac{\partial V}{\partial t}\,dt + \cdots$$

$$= \Delta dS + \frac{1}{2}\Gamma\, dS^2 + \Theta\, dt + \cdots \quad (9.9)$$

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 price-yield relationship is now

$$dV = -(D^*V)\,dy + \frac{1}{2}(CV)\,dy^2 + \cdots \tag{9.10}$$

where the second-order coefficient C is called *convexity* and is akin to Γ.

Figure 9–4 describes the approximation for a simple position, a long position in a European call option. It shows that the linear model is valid only for small movements around the initial value. For larger movements, the delta-gamma approximation creates a better fit.

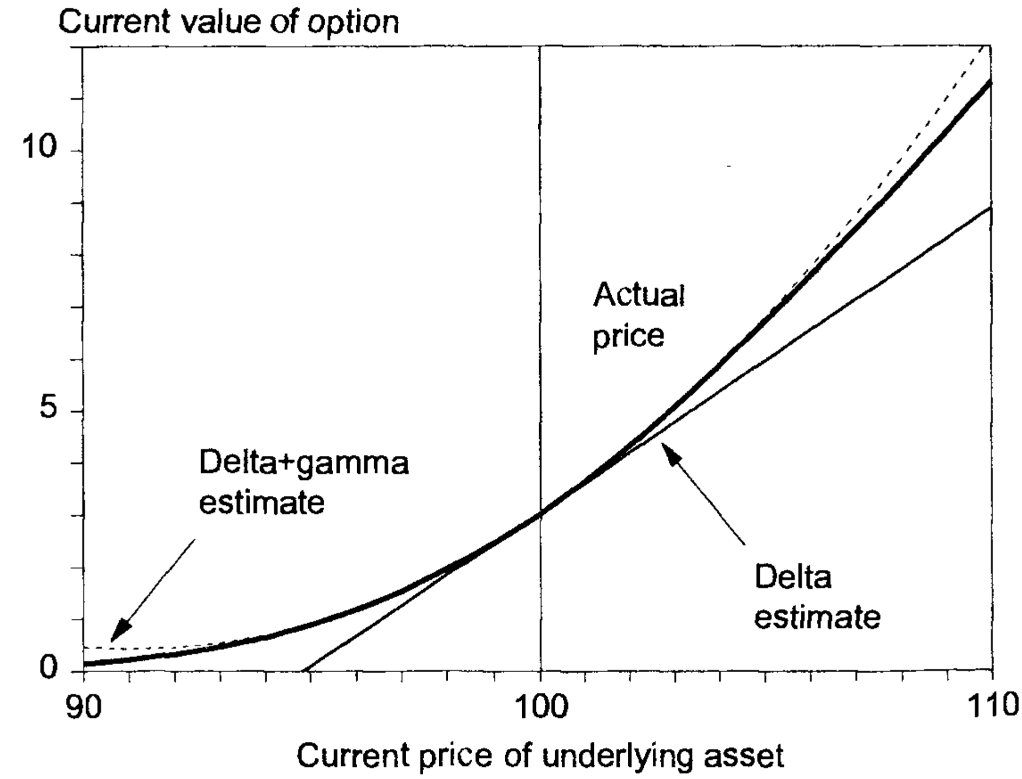
We use the Taylor expansion to compute VAR for the long-call option in Equation (9.6), which yields

$$\begin{aligned}\text{VAR} &= V(S_0) - V(S_0 - \alpha\sigma S_0)\\ &= V(S_0) - [V(S_0) + \Delta(-\alpha\sigma S) + 1/2\Gamma(-\alpha\sigma S)^2]\\ &= |\Delta|(\alpha\sigma S) - 1/2\Gamma(\alpha\sigma S)^2\end{aligned} \tag{9.11}$$

This formula is actually valid for long and short positions in calls and puts. If Γ is positive, which corresponds to a net long position in options, the second term will decrease the linear VAR. Indeed, Figure 9–4

FIGURE 9–4

Delta-gamma approximation for a long call.



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 transformation does not apply, unfortunately, to more complex functions $V(S)$, so we have to go back to the Taylor expansion [Equation (9.9)]. The question now is how to deal with the random variables dS and dS^2.

The simplest method is called the *delta-gamma-delta method.* Taking the variance of both sides of the quadratic approximation [Equation (9.9)], we obtain

$$\sigma^2(dV) = \Delta^2\sigma^2(dS) + (1/2\Gamma)^2\sigma^2(dS^2) + 2(\Delta 1/2\Gamma)\ \mathrm{cov}(dS, dS^2) \quad (9.12)$$

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

$$\sigma^2(dV) = \Delta^2\sigma^2(dS) + 1/2[\Gamma\sigma^2(dS)]^2 \quad (9.13)$$

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

$$\mathrm{VAR} = \alpha\sqrt{(\Delta S\sigma)^2 + 1/2(\Gamma S^2\sigma^2)^2} \quad (9.14)$$

This is, of course, only an approximation. Even if dS was normal, its square dS^2 could not possibly also 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.[2] The corrected VAR, using the so-called *Cornish-Fisher expansion,* is then obtained by replacing α in Equation (9.14) by

$$\alpha' = \alpha - \tfrac{1}{6}(\alpha^2 - 1)\xi \quad (9.15)$$

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.[3]

The second method is the *delta-gamma–Monte Carlo method,* which creates random simulations of the risk factors S and then uses the Taylor

2. Skewness can be computed as $\xi = [E(dV^3) - 3E(dV^2)E(dV) + 2E(dV)^3]/\sigma^3(dV)$ using the third moment of dV, which is $E(dV^3) = (9/2)\Delta^2\Gamma S^4\sigma^4 + (15/8)\Gamma^3 S^6\sigma^6$.

3. See also Zangari (1996).

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 portfolio is fully valued at the initial point V_0 only. The VAR can then be found from the empirical distribution of the portfolio value.

In theory, the delta-gamma method could be generalized to many sources of risk. In a multivariate framework, the Taylor expansion is

$$dV(S) = \Delta' dS + 1/2(dS)'\Gamma(dS) + \cdots \tag{9.16}$$

where dS is now a vector of N changes in market prices, Δ is a vector of N deltas, and Γ is an N by N symmetrical matrix of gammas with respect to the various risk factors. While the diagonal components are conventional gamma measures, the off-diagonal terms are *cross-gammas*, or $\Gamma_{i,j} = \partial^2 V/\partial S_i \partial S_j$. For instance, the delta of options also depends on the implied volatility, which creates a cross-effect.

Unfortunately, the delta-gamma method is not practical with many sources of risk because the amount of data required increases geometrically. For instance, with $N = 100$, we need 100 estimates of Δ, 5050 estimates for the covariance matrix Σ, and an additional 5050 for the matrix Γ, which includes second derivatives of each position with respect to each source of risk. In practice, only the diagonal components are considered. Even so, a full Monte Carlo method provides a more direct route to VAR measurement for large portfolios.

9.1.4 Comparison of Methods

To summarize, Table 9–1 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 portfolios exposed to a few sources of risk and with substantial option components, the "Greeks" method provides increased precision at a low computational cost.
- For portfolios with substantial option components (such as mortgages) or longer horizons, a full-valuation method may be required.

It should be noted that the linear/nonlinear dichotomy also has implications for the choice of the VAR horizon. With linear models, as w

TABLE 9-1

Comparison of VAR Methods

Risk Factor Distribution	Valuation Method	
	Local Valuation	**Full Valuation**
Analytical	Delta-normal Delta-gamma-delta	Not used
Simulated	Delta-gamma-MC	Monte Carlo (MC) Grid MC Historical

have seen in Chapter 4, daily VAR can be adjusted easily to other periods by simple scaling by a square root of time factor. This adjustment assumes that the position is constant and that daily returns are independent and identically distributed.

This time adjustment, however, is not valid for options positions. Since options can be replicated by dynamically changing positions in the underlying assets, the risk of options positions can be dramatically different from the scaled measure of daily risk. Therefore, *adjustments of daily volatility to longer horizons using the square root of time factor are valid only when positions are constant and when optionality in the portfolio is negligible.* For portfolios with substantial options components, the full-valuation method must be implemented over the desired horizon instead of scaling a daily VAR measure.

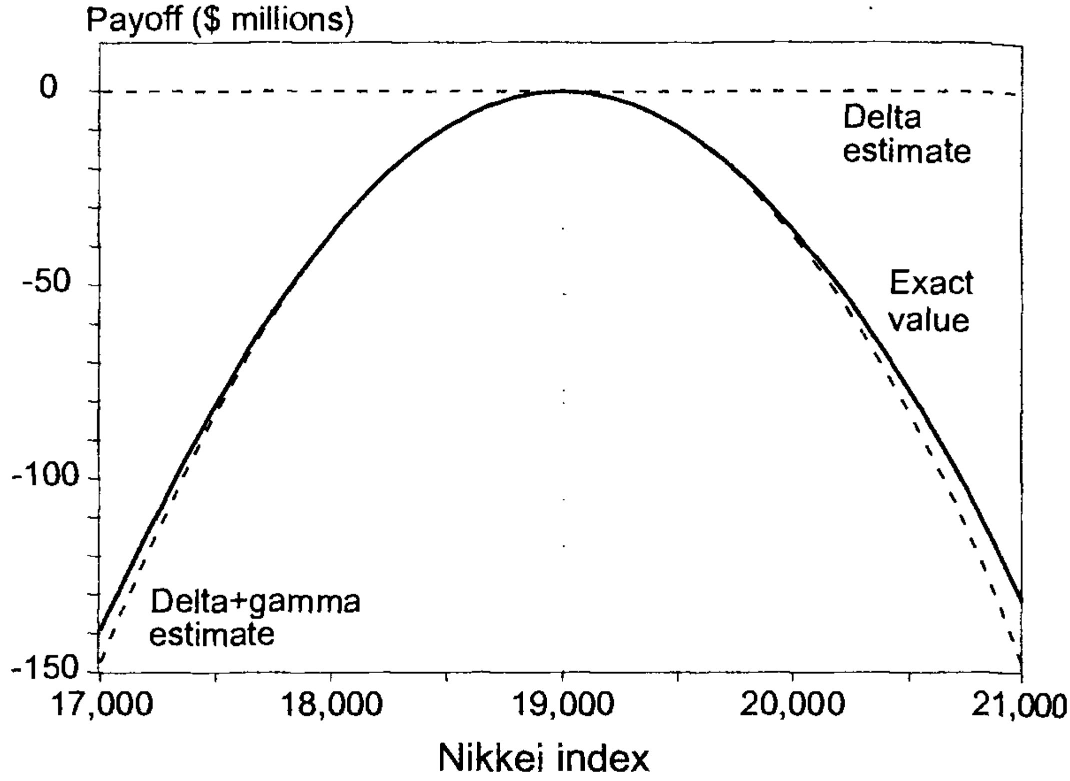
9.1.5 An Example: Leeson's Straddle

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 most of the options were at-the-money.

Leeson did not deal in small amounts. With a multiplier of 500 yen for the options contract and a 100-yen/$ exchange rate, the dollar exposure of the call options to the Nikkei was delta times $0.175 million. Initially, the market value of the position was zero. The position was

FIGURE 9–5

Leeson's straddle.



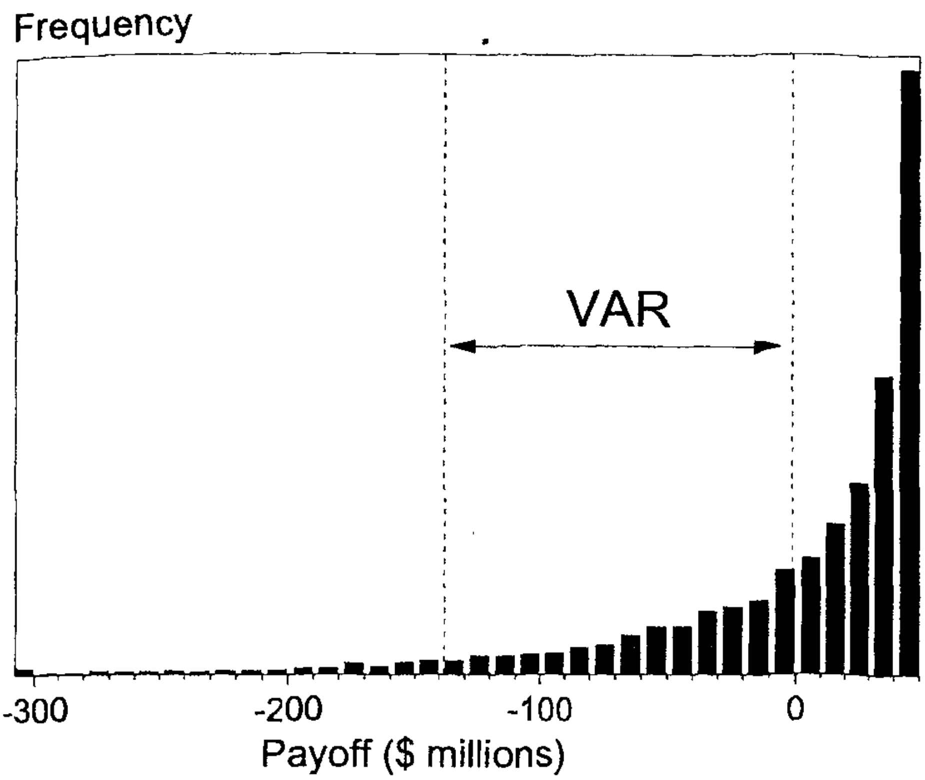
designed to turn in a profit if the Nikkei remained stable. Unfortunately, it also had an unlimited potential for large losses.

Figure 9–5 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. How do we compute the potential loss over a horizon of, say, 1 month?

The risks involved are described in Figure 9–6, 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

FIGURE 9–6

Distribution of 1-month payoff for straddle.



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.

How does the "Greeks" method fare for this portfolio? 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 × 0.000422 = \$0.0000739 million. Over a 1-month horizon, the standard deviation of the Nikkei is $\sigma S = 19{,}000 \times 20$ percent/$\sqrt{12} = 1089$. Ignoring the time drift, the VAR is, from Equation (9.13),

$$\text{VAR} = \alpha\sqrt{\tfrac{1}{2}[\Gamma(\sigma S)^2]^2} = 1.65\sqrt{\tfrac{1}{2}(\$0.0000739 \text{ million} \times 1089^2)^2}$$
$$= 1.65 \times \$62 \text{ million} = \$102 \text{ million}$$

This is substantially better than the delta-normal VAR of zero, which could have fooled us into believing the position was riskless.

Using the Cornish-Fisher expansion and a skewness coefficient of −2.83, we obtain a correction factor of $\alpha' = 1.65 - \frac{1}{6}(1.65^2 - 1)(-2.83)$

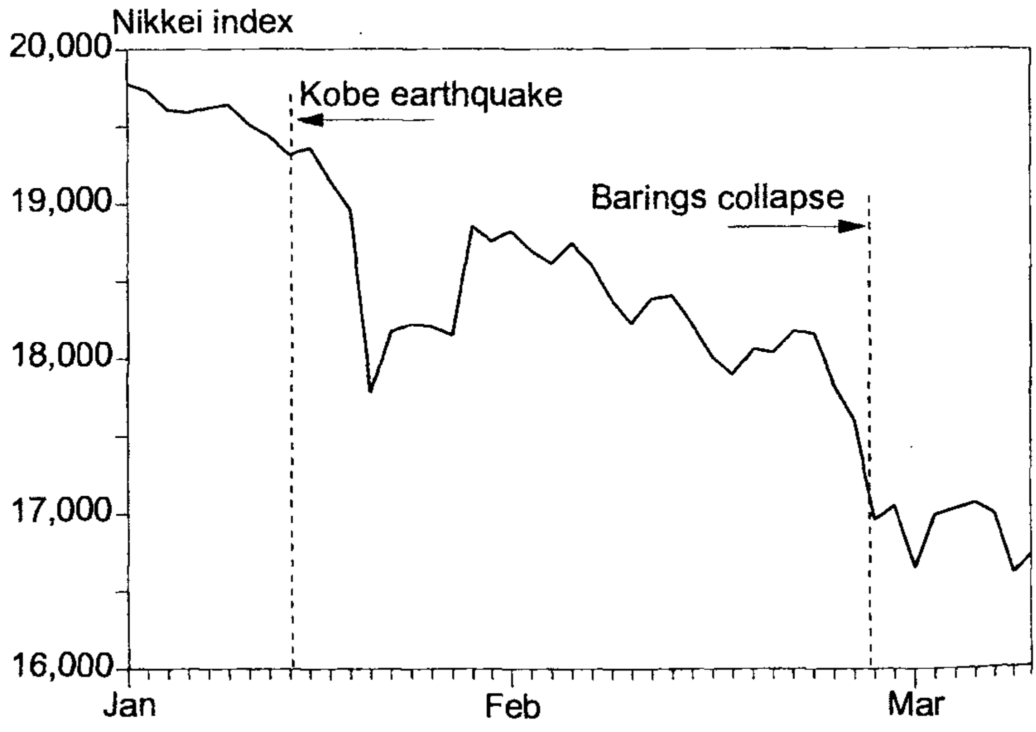
= 2.45. The refined VAR measure is then 2.45 × \$62 million = \$152 million, much closer to the true value of \$138 million.

Finally, we can turn to 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. This variety of methods shows that the straddle had substantial downside risk.

And indeed the options 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 options 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 9–7. 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 9–7

The Nikkei's fall.



9.2 DELTA-NORMAL METHOD

9.2.1 Implementation

If the portfolio consisted of only securities with jointly normal distributions, the measurement of VAR would be relatively simple. The portfolio return is

$$R_{p,t+1} = \sum_{i=1}^{N} w_{i,t} R_{i,t+1} \tag{9.17}$$

where the weights $w_{i,t}$ are indexed by time to recognize the dynamic nature of trading portfolios.

Since the portfolio return is a linear combination of normal variables, it is also normally distributed. Using matrix notations, the portfolio variance is given by

$$\sigma^2(R_{p,t+1}) = w_t' \Sigma_{t+1} w_t \tag{9.18}$$

where Σ_{t+1} is the forecast of the covariance matrix over the VAR horizon.

The problem is that VAR must be measured for large and complex portfolios that evolve over time. The delta-normal method, which is explained in much greater detail in a subsequent chapter, simplifies the process by

- Specifying a list of risk factors
- Mapping the linear exposure of all instruments in the portfolio onto these risk factors
- Aggregating these exposures across instruments
- Estimating the covariance matrix of the risk factors
- Computing the total portfolio risk

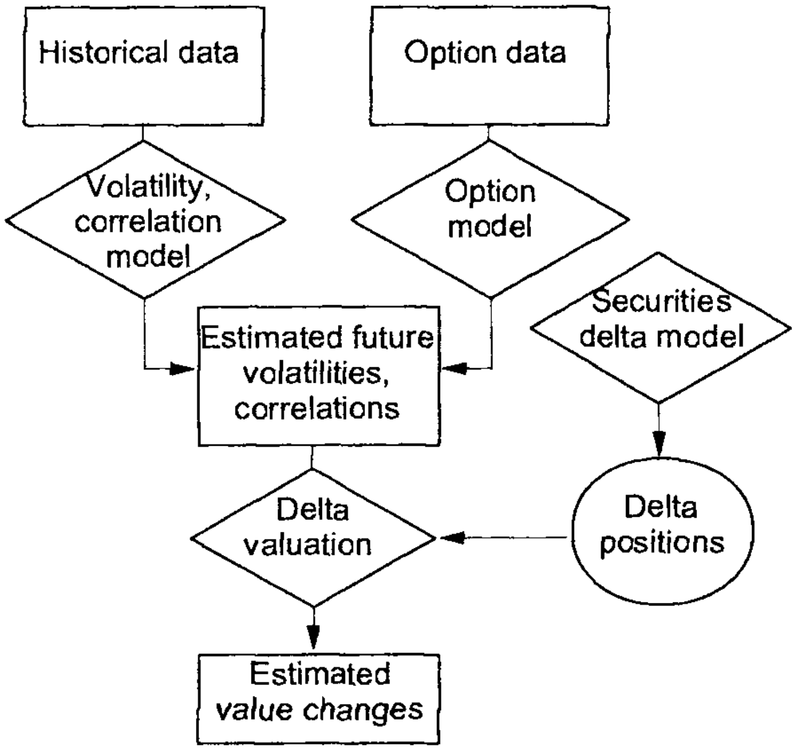
This mapping produces a set of exposures $x_{i,t}$ aggregated across all instruments for each risk factor and measured in dollars. The portfolio VAR is then

$$\text{VAR} = \alpha \sqrt{x_t' \Sigma_{t+1} x_t} \tag{9.19}$$

Within this class of models, two methods can be used to measure the variance-covariance matrix Σ. It can be solely based on historical data using, for example, a model that allows for time variation in risk. Alternatively, it can include implied risk measures from options. Or it can

FIGURE 9–8

Delta-normal method.



use a combination of both. As we saw in the preceding chapter, options-implied measures of risk are superior to historical data but are not available for every asset, let alone for pairs of assets. Figure 9–8 details the steps involved in this approach.

9.2.2 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 parametric approach, VAR is easily *amenable to analysis*, since measures of marginal and incremental risk are a by-product of the VAR computation.

9.2.3 Problems

The delta-normal method can be subject to a number of criticisms. 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 value at risk. As discussed in Chapter 8, some of these fat tails can be explained in terms of time variation in risk. However, even after adjustment, there are still too many observations in the tails. A simple ad hoc adjustment consists of increasing the parameter α to compensate, as is explained in Chapter 5.

Another problem is that the method inadequately measures the risk of *nonlinear instruments*, such as options or mortgages. Under the delta-normal method, options positions are represented by their "deltas" relative to the underlying asset. As we have seen in the preceding section, asymmetries in the distribution of options are not captured by the delta-normal VAR.

Lest we lead you into thinking that this method is inferior, we will now show that alternative methods are no panacea because they involve a quantum leap in difficulty. The delta-normal method is computationally easy to implement. It only requires the market values and exposures of current positions, combined with risk data. Also, in many situations, the delta-normal method provides adequate measurement of market risks.

9.3 HISTORICAL SIMULATION METHOD

9.3.1 Implementation

The historical simulation method provides a straightforward implementation of full valuation (Figure 9–9). It consists of going back in time, such as over the last 250 days, and applying current weights to a time-series of historical asset returns:

$$R_{p,k} = \sum_{i=1}^{N} w_{i,t} R_{i,k} \qquad k = 1, \ldots, t \tag{9.20}$$

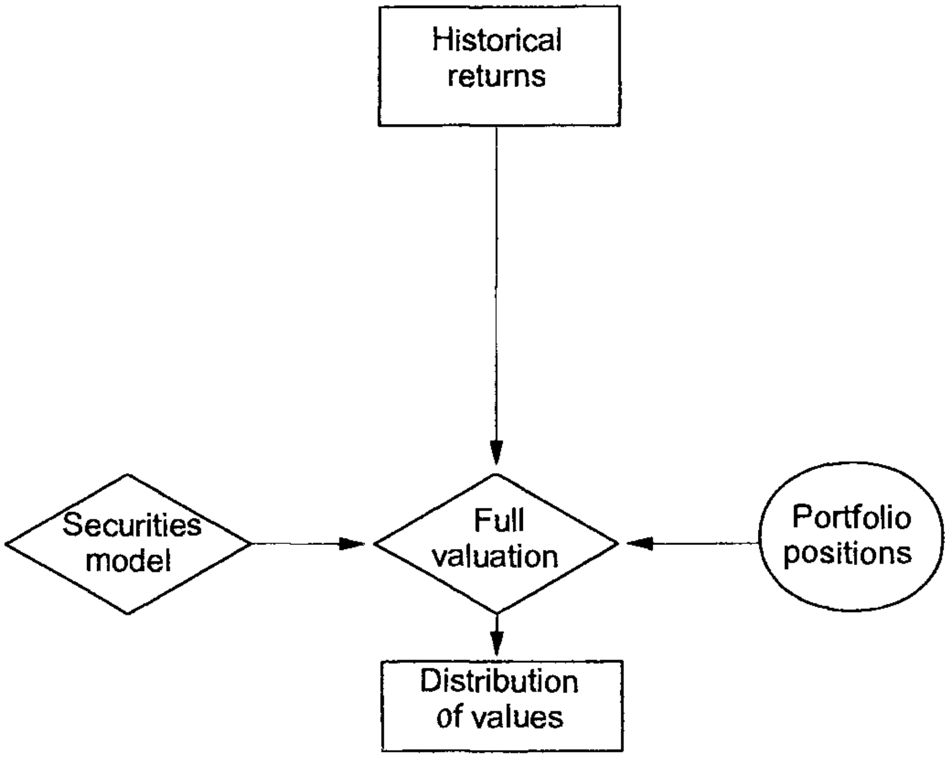
Note that the weights w_t are kept at their current values. This return does not represent an actual portfolio but rather reconstructs the history of a hypothetical portfolio using the current position. The approach is sometimes called *bootstrapping* because it involves using the actual distribution of recent historical data (without replacement).

More generally, full valuation requires a set of complete prices, such as yield curves, instead of just returns. Hypothetical future prices for scenario k are obtained from applying historical changes in prices to the current level of prices:

$$S^*_{i,k} = S_{i,0} + \Delta S_{i,k} \qquad i = 1, \ldots, N \tag{9.21}$$

FIGURE 9–9

Historical simulation method.



A new portfolio value $V^*_{p,k}$ is then computed from the full set of hypothetical prices, perhaps incorporating nonlinear relationships $V^*_k = V(S^*_{i,k})$. Note that to capture *vega risk*, due to changing volatilities, the set of prices can incorporate implied volatility measures. This creates the hypothetical return corresponding to simulation k:

$$R_{p,k} = \frac{V^*_k - V_0}{V_0} \tag{9.22}$$

VAR is then obtained from the entire distribution of hypothetical returns, where each historical scenario is assigned the same weight of $(1/t)$.

As always, the choice of the sample period reflects a tradeoff between using longer and shorter sample sizes. Longer intervals increase the accuracy of estimates but could use irrelevant data, thereby missing important changes in the underlying process.

9.3.2 Advantages

This method is relatively *simple to implement* if historical data have been collected in-house for daily marking-to-market. The same data can then 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 return.

The method also deals directly with the *choice of horizon* for measuring VAR. Returns are simply 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.

By relying on actual prices, the method allows nonlinearities and nonnormal distributions. *Full valuation* is obtained in the simplest fashion: from historical data. The method captures gamma, vega risk, and correlations. It does not rely on specific assumptions about valuation models or the underlying stochastic structure of the market.

Perhaps most important, it can account for *fat tails* and, because it does not rely on valuation models, is not prone to model risk. The method is robust and intuitive and, as such, is perhaps the most widely used method to compute VAR.

9.3.3 Problems

On the other hand, the historical simulation method has a number of drawbacks. First, it assumes that we do have a *sufficient history* of price changes. To obtain 1000 independent simulations of a 1-day move, we require 4 years of continuous data. Some assets may have short histories, or there may not be a record of an asset's history.

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.

And as we have demonstrated in Chapter 8, risk contains significant and predictable time variation. The simple historical simulation method presented here will miss situations with temporarily elevated volatility.[4] Worse, historical simulation will be very slow to incorporate *structural breaks*, which are handled more easily with an analytical methods such as RiskMetrics.

---

4. A simple method to allow time variation in risk proceeds as follows: First, fit a time-series model to the conditional volatility and construct historical scaled residuals. Second, perform a historical simulation on these residuals. Third, apply the most recent volatility forecast to the scaled portfolio volatility. For applications, see Hull and White (1998).

This approach is also subject to the same criticisms as the *moving-window estimation of variances*. The method puts the same weight on all observations in the window, including old data points. The measure of risk can change significantly after an old observation is dropped from the window.[5]

Likewise, the *sampling variation* of the historical simulation VAR will be much greater than for an analytical method. As is pointed out in Chapter 5, VAR is only a statistical estimate and may be subject to much estimation error if the sample size is too short. For instance, a 99 percent daily VAR estimated over a window of 100 days produces only one observation in the tail, which necessarily leads to an imprecise VAR measure. Thus very long sample paths are required to obtain meaningful quantiles. The dilemma is that this may involve observations that are not relevant.

A final drawback is that the method quickly becomes *cumbersome for large portfolios* with complicated structures. In practice, users adopt simplifications such as grouping interest rate payoffs into bands, which considerably increases the speed of computation. Regulators also have adopted such a "bucketing" approach. But if too many simplifications are carried out, such as replacing assets by their delta equivalents, the benefits of full valuation can be lost.

9.4 MONTE CARLO SIMULATION METHOD

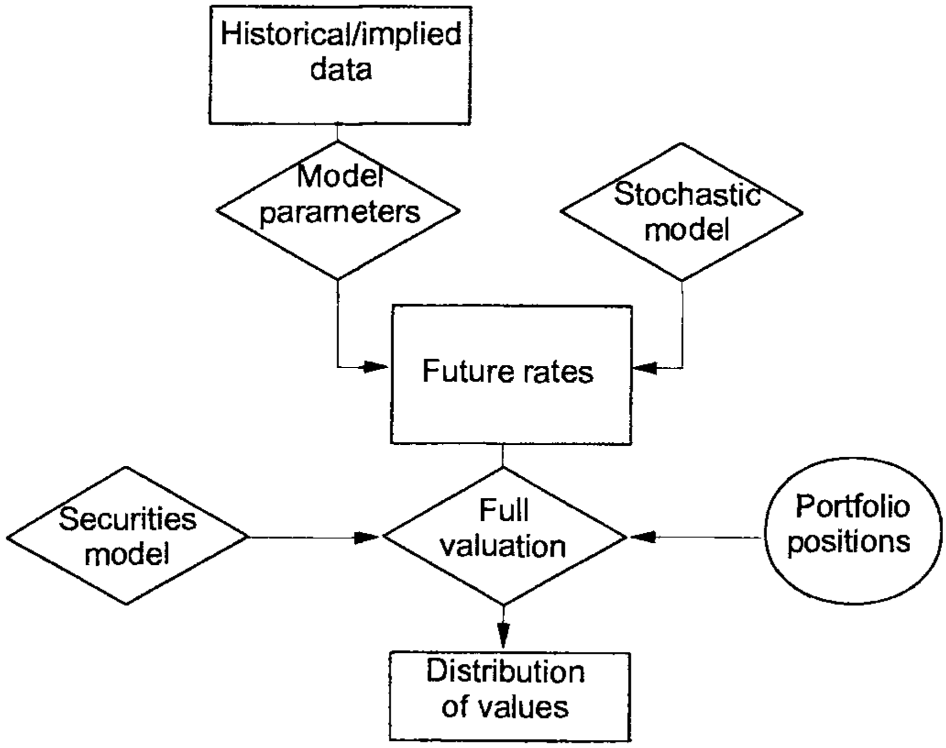
9.4.1 Implementation

Monte Carlo (MC) simulations cover a wide range of possible values in financial variables and fully account for correlations. MC simulation is developed in more detail in a later chapter. In brief, the method proceeds in two steps. First, the risk manager specifies a stochastic process for financial variables as well as process parameters; parameters such as risk and correlations can be derived from historical or options data. Second, fictitious price paths are simulated for all variables of interest. At each horizon considered, the portfolio is marked-to-market using full valuation as in the historical simulation method, $V_k^* = V(S_{i,k}^*)$. Each of these "pseudo" realizations is then used to compile a distribution of returns,

5. To alleviate this problem, Boudoukh et al. (1998) propose a scheme whereby each observation R_k is assigned a weight w_k that declines as it ages. The distribution is then obtained from ranking the R_k and cumulating the associated weights to find the selected confidence level.

FIGURE 9–10

Monte Carlo method.



from which a VAR figure can be measured. The method is summarized in Figure 9–10.

The Monte Carlo method is thus similar to the historical simulation method, except that the hypothetical changes in prices ΔS_i for asset i in Equation (9.20) are created by random draws from a prespecified stochastic process instead of sampled from historical data.

9.4.2 Advantages

Monte Carlo analysis is by far the most *powerful method* to compute VAR. It can account for a wide range of exposures and risks, including *nonlinear price risk,* volatility risk, and even model risk. It is flexible enough to incorporate time variation in volatility, *fat tails,* and extreme scenarios. Simulations generate *the entire pdf,* not just one quantile, and can be used to examine, for instance, the expected loss beyond a particular VAR.

MC simulation also 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; or 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.

9.4.3 Problems

The biggest drawback of 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 itself 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 intellectual development. The MC simulation method is relatively onerous to develop from scratch, despite rapidly falling prices for hardware. 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 as well as pricing models for securities such as options or mortgages. Therefore, it is subject to the risk that the models are wrong. To check if the results are robust to changes in the model, simulation results should be complemented by some sensitivity analysis.

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 will then provide the correct measure of VAR, in one easy step. MC simulations based on the same covariance matrix will give only an approximation, albeit increasingly good as the number of replications increases.

Overall, this method is probably the most comprehensive approach to measuring market risk if modeling is done correctly. To some extent, the method can even handle credit risks. This is why a full chapter is devoted to the implementation of Monte Carlo simulation methods.

9.5 EMPIRICAL COMPARISONS

It is instructive to compare the VAR numbers obtained from the three methods discussed. 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 9–2 summarizes the results, which are compared in terms of percentage of outcomes falling within the VAR forecast. The middle column shows that all methods give a coverage that is very close to the ideal number, which is the 95 percent confidence level. At the 99 percent confidence level, however, the delta-normal methods seem to underestimate VAR slightly, since their coverage falls short of the ideal 99 percent.

Hendricks also reports 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

TABLE 9–2

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%

distribution with a greater α parameter. A Student t distribution with four to six degrees of freedom, for example, would be appropriate.

As important, when the VAR number is exceeded, the tail event is, on average, 30 to 40 percent greater than the risk measure. In some instances, it is several times greater. As Hendricks states, "This makes it clear that VAR measures—even at the 99th percentile—do not bound possible losses."

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 9–3 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 seconds. 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 approximations 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 due to its simplicity. With the rapidly decreasing cost of computing power, however, there is now

TABLE 9–3

Accuracy and Speed of VAR Methods: 99 Percent VAR for Option Portfolios

Method	Accuracy: Mean Absolute Error in VAR (%)	Speed: Computation Time, s
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 Monto Carlo	0	66.27

a marked trend toward the generalized use of historical simulation methods.

9.6 SUMMARY

We can distinguish a number of different methods to measure VAR. At the most fundamental level, they separate into local (or analytical) valuation and full valuation. This separation reflects a tradeoff between speed of computation and accuracy of valuation.

Delta models can use parameters based on historical data, such as those implemented by RiskMetrics, or on implied data, where volatilities are derived from options. Both methods generate a covariance matrix, to which the "delta" or linear positions are applied to find the portfolio VAR. Among full-valuation models, the historical simulation method is the easiest to implement. It simply relies on historical data for securities valuation and applies the most current weight to historical prices. Finally, the most complete model, but also the most difficult to implement, is the Monte Carlo simulation approach, which imposes a particular stochastic process on the financial variables of interest, from which various sample paths are simulated. Full valuation for each sample path generates a distribution of portfolio values.

Table 9–4 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 (nor embedded options) and whose distributions are close to the normal pdf, 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 (due to faulty assumptions or computations). The resulting VAR is easy to explain to management and to the public. Because the method is analytical, it allows easy analysis of the VAR results using marginal and component VAR measures. For portfolios with options positions, however, the method may not be appropriate. Instead, users should turn to a full-valuation method.

The second method, historical simulation, is also relatively easy to implement and uses actual, full valuation of all securities. However, its typical implementation does not account for time variation in risk, and the method relies on a narrow window only.

In theory, the Monte Carlo approach can alleviate all these technical difficulties. It can incorporate nonlinear positions, nonnormal distributions, implied parameters, and even user-defined scenarios. The price to pay for this flexibility, however, is heavy. Computer and data require-

TABLE 9–4

Comparison of Approaches to VAR

Features	Delta-Normal	Historical Simulation	Monto 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	Intermediate	No
Accuracy	Depends on portfolio	Yes	Yes
Communicability	Easy	Easy	Difficult
VAR analysis	Easy, analytical	More difficult	More difficult
Major pitfalls	Nonlinearities, fat tails	Time-variation in risk, unusual events	Model risk

ments are a quantum step above the other two approaches, model risk looms large, and value at risk 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. A recent survey by Britain's Financial Services Authority has revealed that 42 percent of banks use the covariance matrix approach, 31 percent use historical simulation, and 23 percent use the Monte Carlo approach. The delta-normal method, which is the easiest to implement, appears to be the most widespread.

All these methods present some advantages. They are also related. Monte Carlo analysis of linear positions with normal returns, for instance, should yield the same result as the delta-normal method. Perhaps the best lesson from this chapter is to check VAR measures with different methodologies and then to analyze the sources of differences.