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

5.2.3 VAR as Equity Capital

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

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

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

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

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

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

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

TABLE 5-1

Credit Rating and Default Rates

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

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

FIGURE 5-5

Losses at and during the horizon.

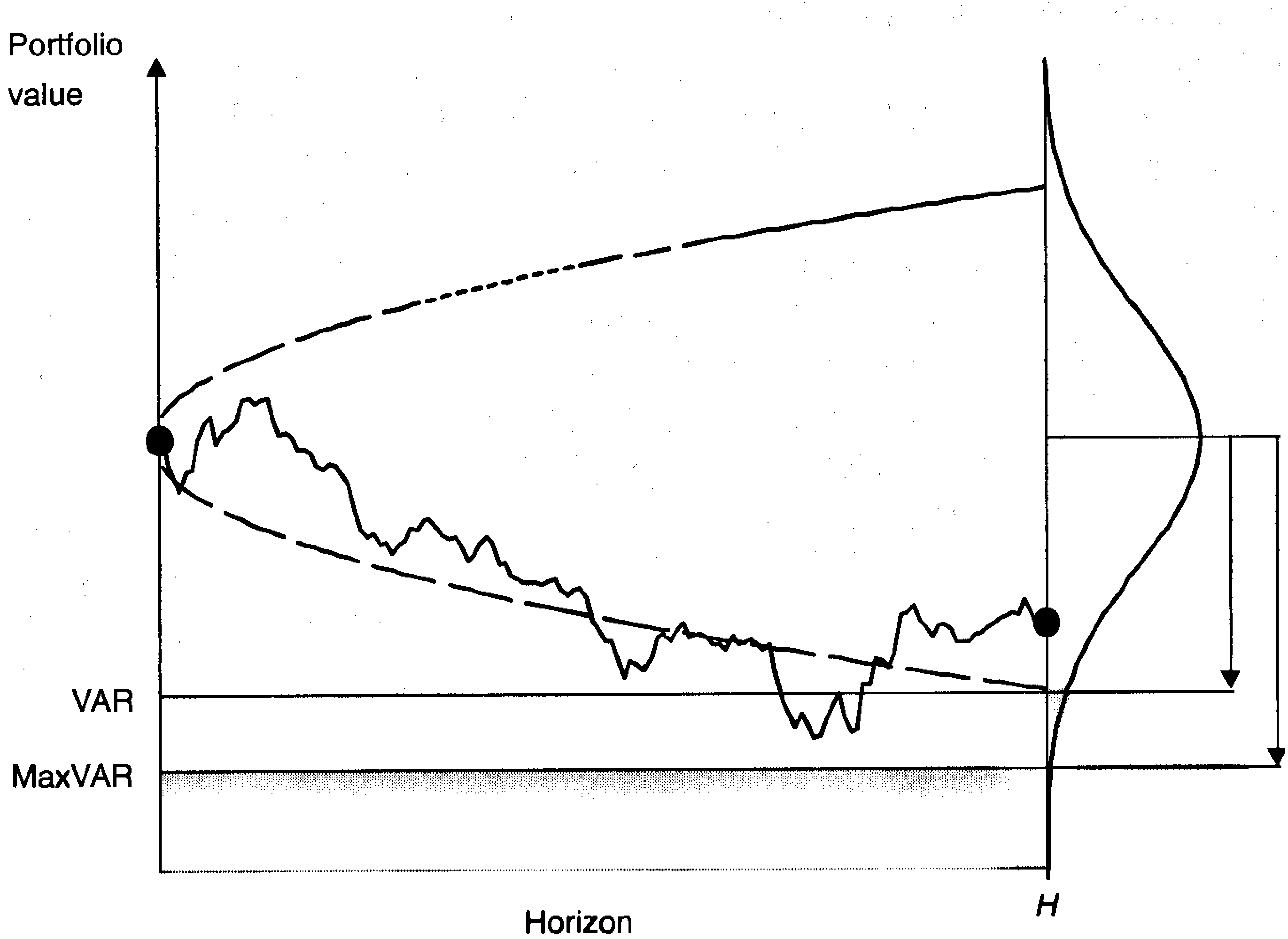


TABLE 5-2

VAR and MaxVAR, Normal Distribution

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

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

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

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

5.2.4 Criteria for Backtesting

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

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

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

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

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

5.2.5 Application: The Basel Parameters

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

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

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

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

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

BOX 5-2

VAR FOR MARGIN REQUIREMENTS

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

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

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

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

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

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

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

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

5.2.6 Conversion of VAR Parameters

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

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

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

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

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

TABLE 5-3

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

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

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

5.3 ASSESSING VAR PRECISION

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

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

5.3.1 The Problem of Measurement Errors

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

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

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

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

5.3.2 Estimation Errors in Means and Variances

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

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

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

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

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

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

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

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

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

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

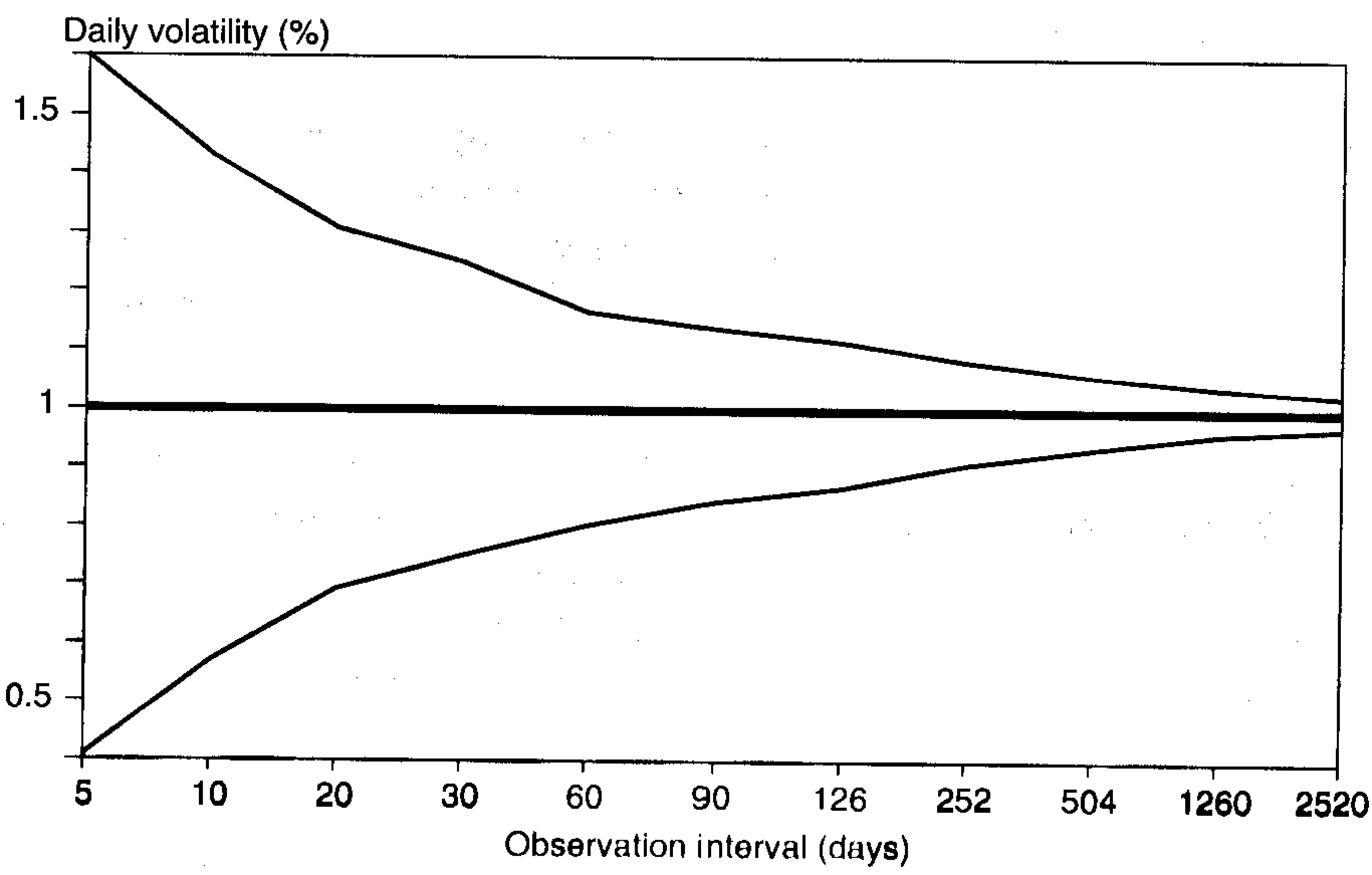
For instance, consider monthly returns on the euro/$ rate from 1973 to 2004. Sample parameters are $\hat{\mu}=-0.15$ percent, $\hat{\sigma}=3.39$ percent, and $T=384$ observations. The standard error of the estimate indicates how confident we are about the sample value; the smaller the error, the more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

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

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

FIGURE 5-6

Confidence bands for sample volatility.



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

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

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

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

5.3.3 Estimation Error in Sample Quantiles

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

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

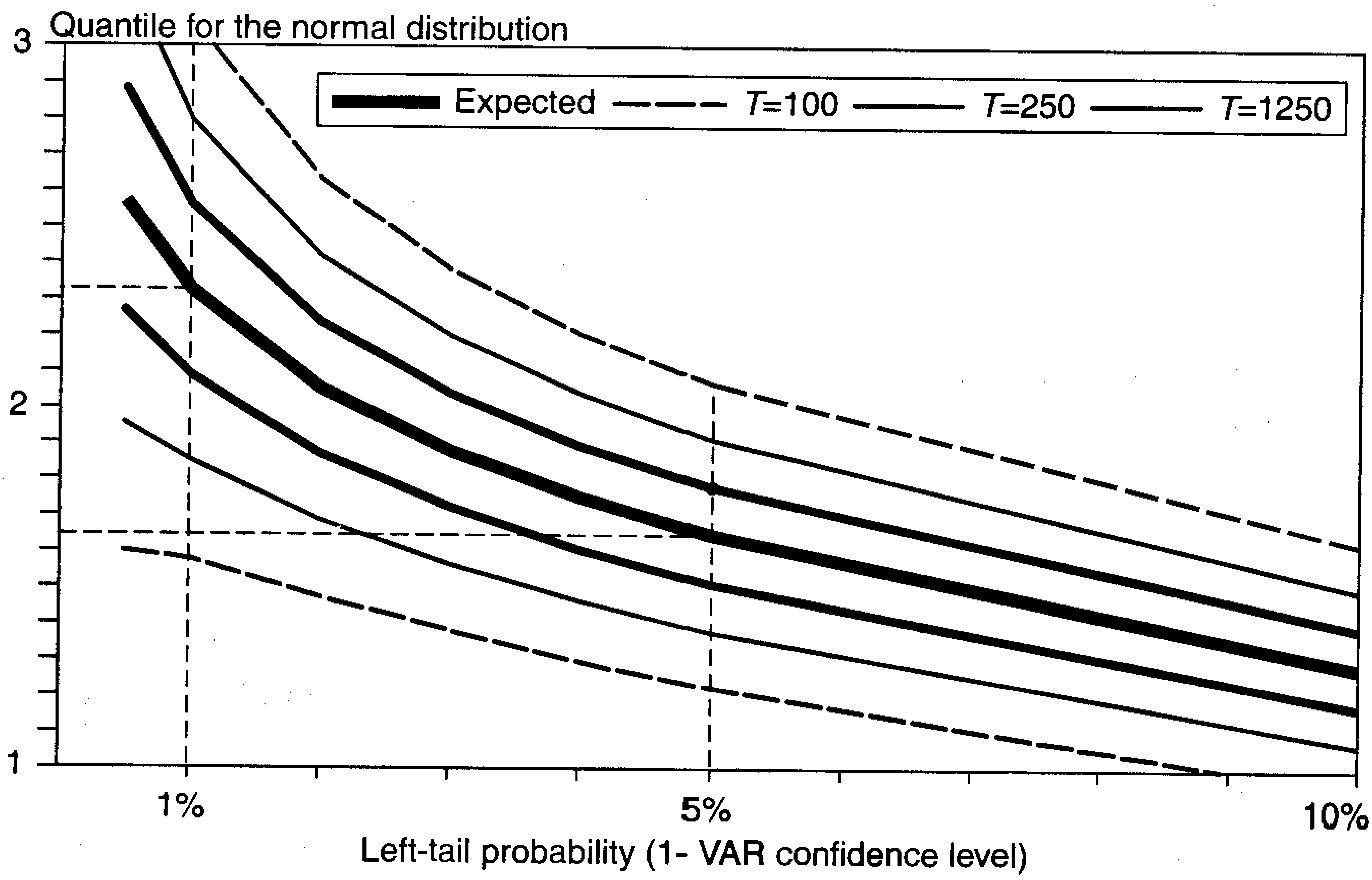
where T is the sample size, and $f(\cdot)$ is the probability distribution function evaluated at the quantile q. The effect of estimation error is illustrated in Figure 5-7, where the expected quantile and 95 percent confidence bands are plotted for quantiles from the normal distribution.

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

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

FIGURE 5-7

Confidence bands for sample quantiles.



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

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

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

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

5.3.4 Comparison of Methods

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

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

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

TABLE 5-4

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

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

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

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

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

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

5.4 EXTREME-VALUE THEORY

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

5.4.1 The EVT Distribution

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

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

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

BOX 5-3

EVT AND NATURAL DISASTERS

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

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

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

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

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

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

5.4.2 Quantiles and ETL

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

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

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

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

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

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

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

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

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

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

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

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

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

FIGURE 5-8

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

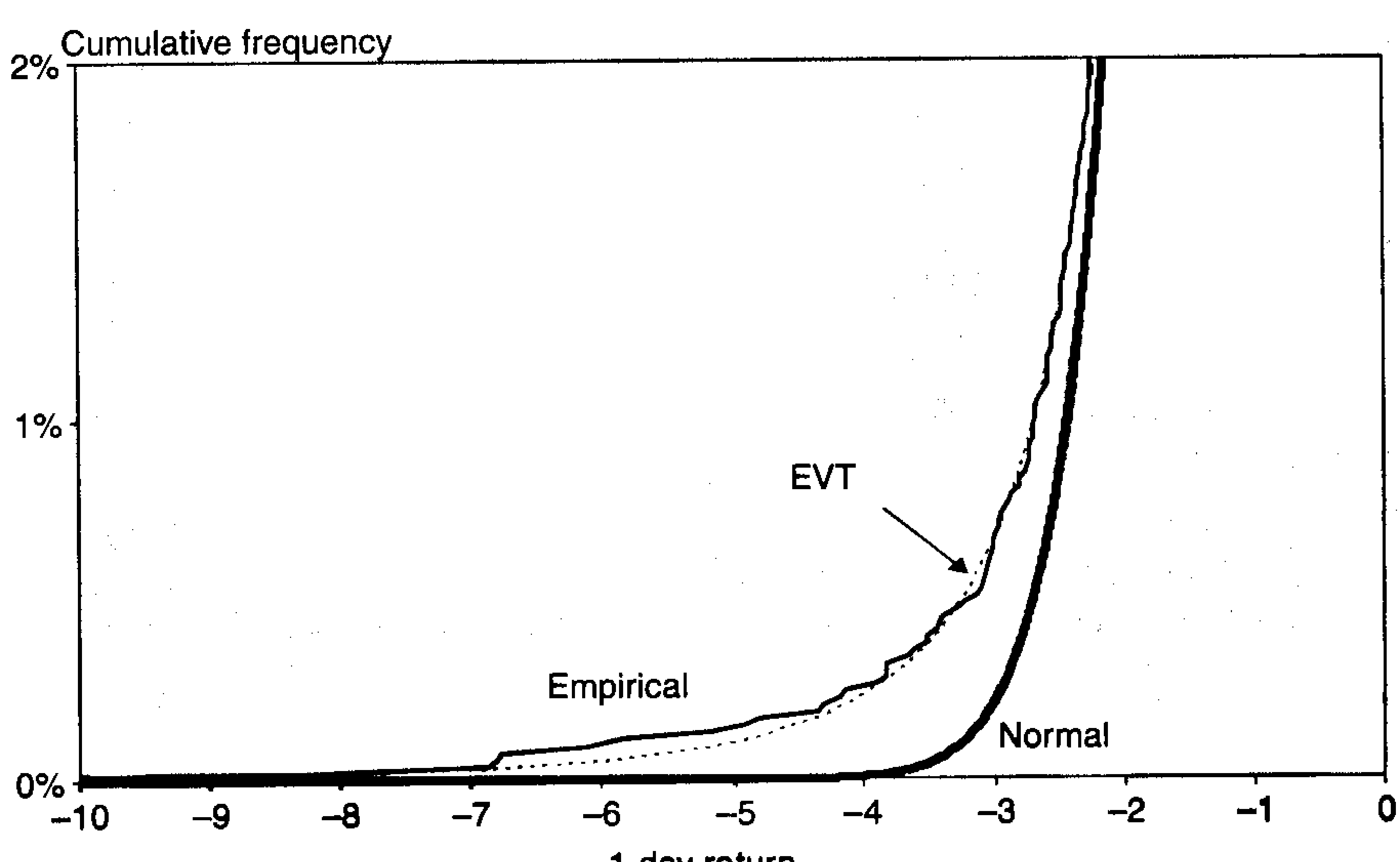


TABLE 5-5

The Effect of Fat Tails and Multiple Periods on VAR

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

Source: Danielsson and de Vries (1997).

5.4.3 Time Aggregation

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

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

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

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

5.4.4 EVT Evaluation

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

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

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

5.5 CONCLUSIONS

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

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

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

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

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

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

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

APPENDIX 5.A

Justification for the Basel Multiplier

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

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

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

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

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

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

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

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

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

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

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

QUESTIONS

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

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

CHAPTER 6

Backtesting VAR

> Disclosure of quantitative measures of market risk, such as value-at-risk, is enlightening only when accompanied by a thorough discussion of how the risk measures were calculated and how they related to actual performance.
>
> —*Alan Greenspan (1996)*

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

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

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

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

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

6.1 SETUP FOR BACKTESTING

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

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

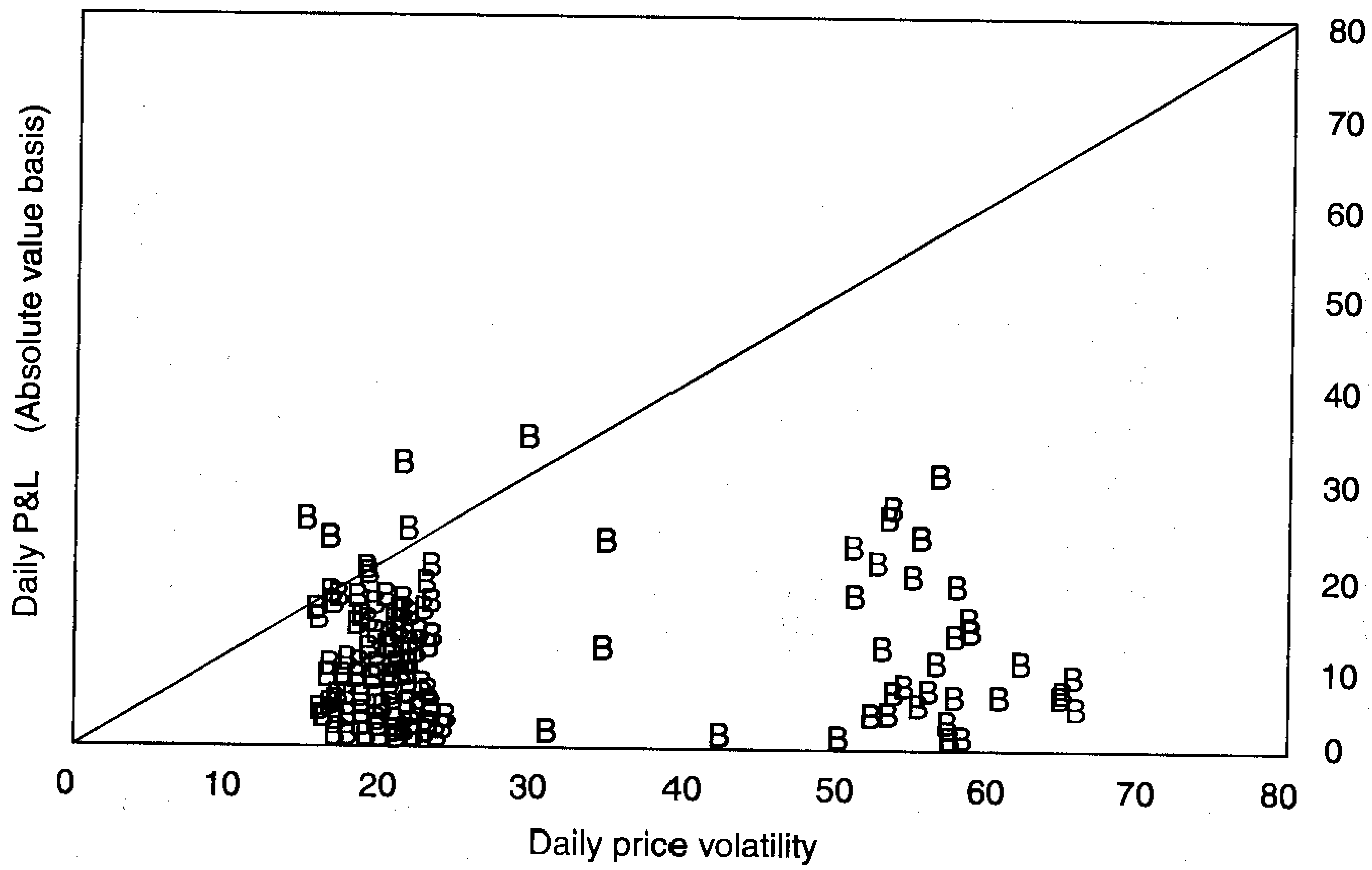
6.1.1. An Example

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

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

FIGURE 6-1

Model evaluation: Bankers Trust.



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

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

6.1.2. Which Return?

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

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

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

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

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

6.2 MODEL BACKTESTING WITH EXCEPTIONS

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

[2] See BCBS (2005b).

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

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

6.2.1. Model Verification Based on Failure Rates

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

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

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

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

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

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

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

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

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

BOX 6-1

J.P. MORGAN'S EXCEPTIONS

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

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

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

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

FIGURE 6-2

Distribution of exceptions when model is correct.

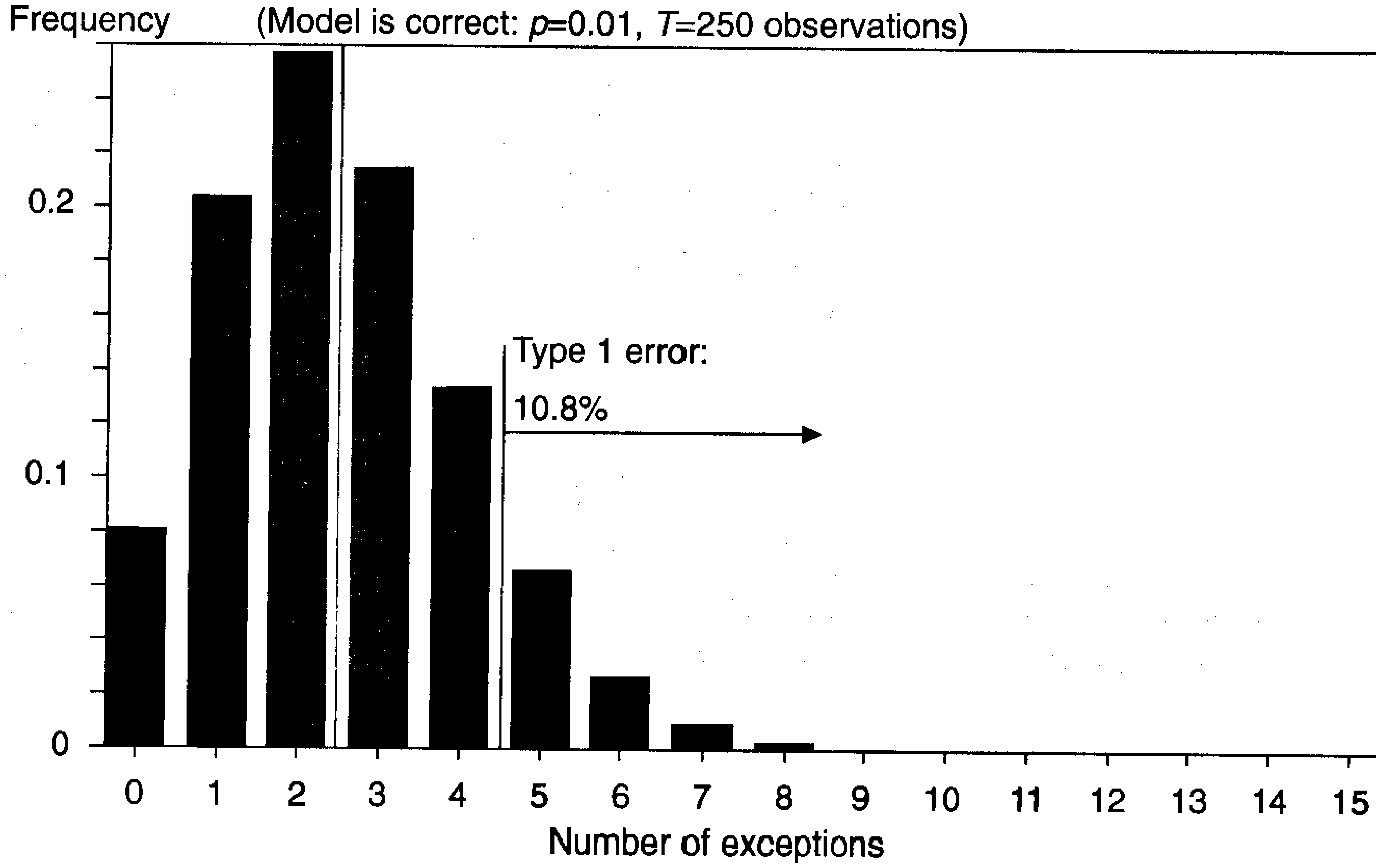
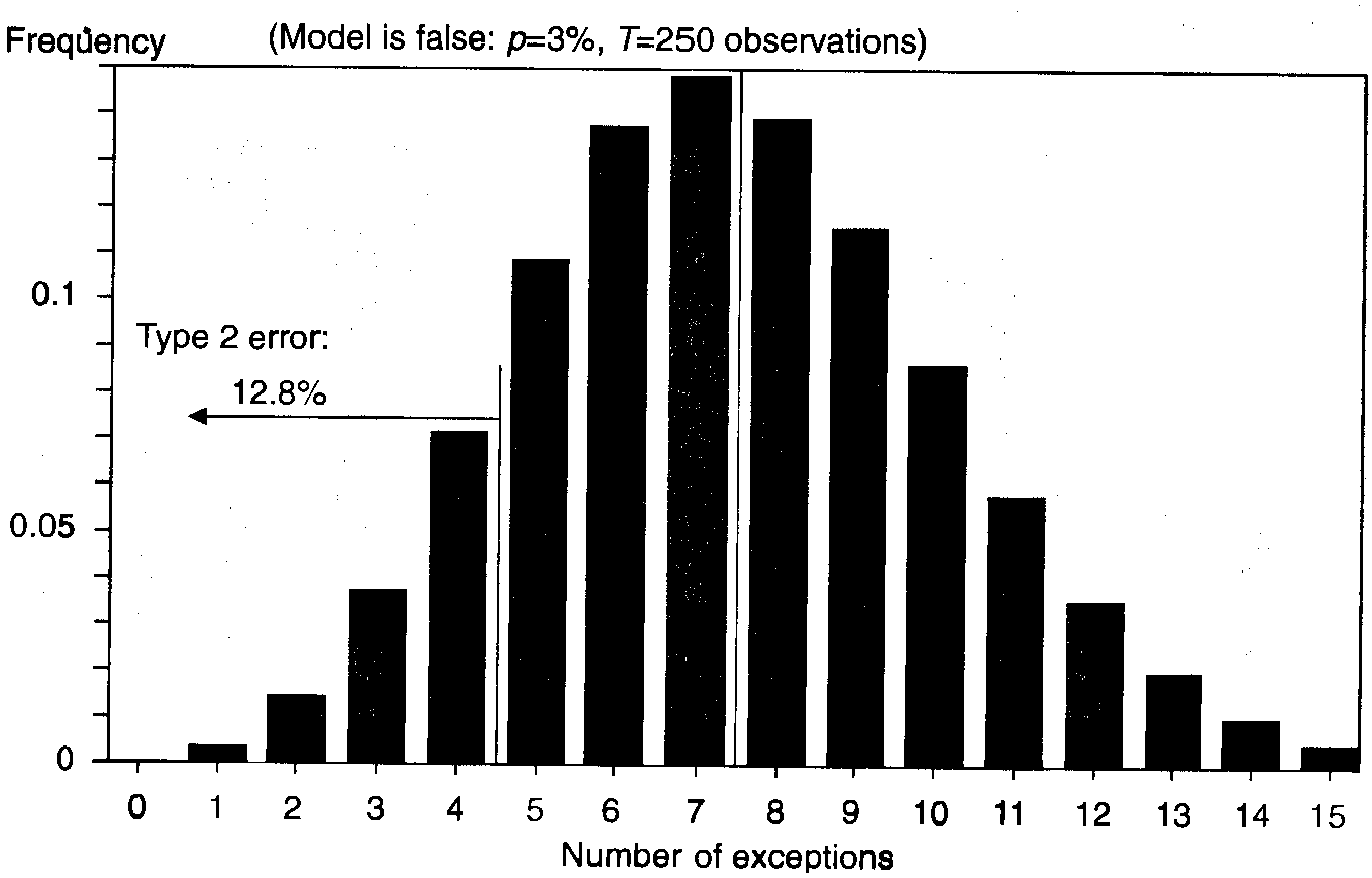


FIGURE 6-3

Distribution of exceptions when model is incorrect.



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

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

TABLE 6-1

Decision Errors

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

TABLE 6-2

Model Backtesting, 95% Nonrejection Test Confidence Regions

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

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

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

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

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

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

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

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

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