TABLE 12-3

Relative Error in 99 Percent VAR for Various Distributions

		Relative Error (Percent)			
		Replications			
Distribution	**Skewness**	**100**	**500**	**1000**	**10,000**
Normal	0.00	17.6	7.3	5.1	1.5
Right skew	0.76	9.3	4.2	3.0	0.9
Left skew	−0.76	23.4	9.2	6.3	1.9

Alternatively, we could search for the number of replications required to measure VAR with a relative error of 1 percent. For the normal distribution, we need more than 20,000 replications to make sure that the relative error in the first row is below 1 percent.

12.3.2 Acceleration Methods

This led to a search for methods to accelerate computations. One of the earliest, and easiest, is the *antithetic variable technique*, which consists of changing the sign of all the random samples ϵ. This method, which is appropriate when the original distribution is symmetric, creates twice the number of replications for the risk factors at little additional cost. We still need, however, twice the original number of full valuations on the target date.

This approach can be applied to the historical simulation method, where we can add a vector of historical price changes with the sign reversed. This is also useful to eliminate the effect of trends in the recent historical data.

Another useful tool is the *control variates technique*. We are trying to estimate VAR, a function of the data sample. Call this $V(X)$. Assume now that the function can be approximated by another function, such as a quadratic approximation $V^Q(X)$, for which we have a closed-form solution v^Q.[7]

[7] This can be found using the anaytical approximations described in Appendix 10.A. Note that v^Q does not depend on the random sample X.

For any sample, the error then is known to be $V^Q(X) - v^Q$ for the quadratic approximation. If this error is highly correlated with the sampling error in $V(X)$, the control variate estimator can be taken as

$$V_{CV} = V(X) - [V^Q(X) - v^Q] \tag{12.5}$$

This estimator has much lower variance than the original one when the quadratic function provides a good approximation of the true function.

The most effective acceleration method is the *importance sampling technique*, which attempts to sample along the paths that are most important to the problem at hand. The idea is that if our goal is to measure a tail quantile accurately, there is no point in doing simulations that will generate observations in the center of the distribution. The method involves shifts in the distribution of random variables. Glasserman et al. (2000) show that relative to the usual Monte Carlo method, the variance of VAR estimators can be reduced by a factor of at least 10.

A related application is the *stratified sampling technique*, which can be explained intuitively as follows: assume that we require VAR for a long position in a call option.[8] We are trying to keep the number of replications at $K = 1000$. To increase the accuracy of the VAR estimator, we could partition the simulation region into two zones. As before, we start from a uniform distribution, which then is transformed into a normal distribution for the underlying asset price using the *inverse transform method.*

Define these two zones, or strata, for the uniform distribution as [0.0, 0.1] and [0.1, 1.0]. Thus *stratification* is the process of grouping the data into mutually exclusive and collectively exhaustive regions. Usually, the probabilities of the random number falling in both zones are selected as $p_1 = 10$ percent and $p_2 = 90$ percent, respectively. Now we change these probabilities to 50 percent for both regions. The number of observations now is $K_1 = 500$ for the first region and $K_2 = 500$ for the second. This increases the number of samples for the risk factor in the first, left-tail region.

Estimators for the mean need to be adjusted for the stratification. We weight the estimator for each region by its probability, that is,

$$E(F_T) = p_1 \frac{\sum_{i=1}^{K_1} F_i}{K_1} + p_2 \frac{\sum_{i=1}^{K_2} F_i}{K_2} \tag{12.6}$$

[8] More generally, the classification can be based on a quadratic approximation. See Cardenas et al. (1999) for more details.

To compute VAR, we simply examine the first region. The 50 percent quantile for the first region, for example, provides an estimator of a $10 \times 0.5 = 5$ percent left-tail VAR. Because VAR only uses the number of observations in the right region, we do not even need to compute their value, which economizes on the time required for full valuation.

This reflects the general principle that using more information about the portfolio distribution results in more efficient simulations. In general, unfortunately, the payoff function is not known. All is not lost, however. Instead, the simulation can proceed in two passes. The first pass runs a traditional Monte Carlo. The risk manager then examines the region of the risk factors that causes losses around VAR. A second pass then is performed with many more samples from this region.

12.4 SIMULATIONS WITH MULTIPLE VARIABLES

Modern risk measurement applications are large-scale problems because they apply at the highest level of the financial institution. This requires simulations over multiple sources of risk.

12.4.1 From Independent to Correlated Variables

Simulations generate independent random variables that need to be transformed to account for correlations. Define N as the number of sources of risk. If the variables are uncorrelated, the randomization can be performed independently for each variable, that is,

$$\Delta S_{j,t} = S_{j,t-1} \left(\mu_j \Delta t + \sigma_j \epsilon_{j,t} \sqrt{\Delta t}\right) \tag{12.7}$$

where the ϵ values are independent across time period and series $j = 1, \ldots, N$.

To account for correlations between variables, we start with a set of independent variables η, which then are transformed into the ϵ. In a two-variable setting, we construct

$$\begin{aligned} \epsilon_1 &= \eta_1 \\ \epsilon_2 &= \rho\eta_1 + (1 - \rho^2)^{1/2}\eta_2 \end{aligned} \tag{12.8}$$

where ρ is the correlation coefficient between the variables ϵ. First, we verify that the variance of ϵ_2 is unity, that is,

$$V(\epsilon_2) = \rho^2 V(\eta_1) + [(1 - \rho^2)^{1/2}]^2 V(\eta_2) = \rho^2 + (1 - \rho^2) = 1$$

Then we compute the covariance of the ϵ *as*

$$\text{cov}(\epsilon_1, \epsilon_2) = \text{cov}[\eta_1, \rho\eta_1 + (1 - \rho^2)^{1/2}\eta_2] = \rho\,\text{cov}(\eta_1, \eta_1) = \rho$$

This confirms that the ϵ variables have correlation of ρ. The question is, how was the transformation in Equation (12.8) chosen?

More generally, suppose that we have a vector of N values of ϵ for which we would like to display some correlation structure $V(\epsilon) = E(\epsilon\epsilon')$ = R. We will use *Cholesky factorization,* named after the French mathematician André-Louis Cholesky, to generate correlated variables. Since the matrix R is a symmetric real matrix, it can be decomposed into its *Cholesky factors*, this is,

$$R = TT' \tag{12.9}$$

where T is a lower triangular matrix with zeros in the upper right corners.

We start with an N vector η that is composed of independent variables all with unit variances. In other words, $V(\eta) = I$, where I is the identity matrix with zeroes everywhere except on the diagonal. Next, construct the variable $\epsilon = T\eta$. Its covariance matrix is $V(\epsilon) = E(\epsilon\epsilon') = E(T\eta\eta'T') = TE(\eta\eta')T' = TIT' = TT' = R$. Thus we have confirmed that the values of ϵ have the desired correlations.

As an example, consider the two-variable case. The matrix can be decomposed into

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^2 & a_{11}a_{12} \\ a_{11}a_{12} & a_{12}^2 + a_{22}^2 \end{bmatrix}$$

The entries in the right-hand-side matrix must match exactly each entry in the correlation matrix. Because the Cholesky matrix is triangular, the factors can be found by successive substitution by setting

$$a_{11}^2 = 1$$
$$a_{11}a_{12} = \rho$$
$$a_{12}^2 + a_{22}^2 = 1$$

which yields $$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & (1-\rho^2)^{1/2} \end{bmatrix}$$

And indeed, this is how Equation (12.8) was obtained:

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{1/2} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$$

This explains how a multivariate set of random variables can be created from simple building blocks consisting of i.i.d. variables. In addition to providing a method to generate correlated variables, however, this approach generates valuable insight into the random number generation process.

12.4.2 Number of Risk Factors

For the decomposition to work, the matrix R must be *positive definite.* Otherwise, there is no way to transform N independent source of risks into N correlated variables of ϵ.

As discussed in Chapter 8, this condition can be verified with the *singular value decomposition.* This decomposition of the covariance matrix provides a check that the matrix is well behaved. If any of the eigenvalues is zero or less than zero, the Cholesky decomposition will fail.

When the matrix R is not positive definite, its *determinant* is zero. Intuitively speaking, the determinant d is a measure of the "volume" of a matrix. If d is zero, the dimension of the matrix is less than N. The determinant can be computed easily from the Cholesky decomposition. Since the matrix T has zeros above its diagonal, its determinant reduces to the product of all diagonal coefficients $d_T = \Pi_{i=1}^{N} a_{ii}$. The determinant of the covariance matrix R then is $d = d_T^2$.

In our two-factor example, the matrix is not positive definite if $\rho = 1$. In practice, this implies that the two factors are really the same. The Cholesky decomposition then yields $a_{11} = 1$, $a_{12} = 1$, and $a_{22} = 0$, and the determinant $d = (a_{11}a_{22})^2$ is 0. As a result, the second factor η_2 is never used, and ϵ_1 is always the same as ϵ_2. The second random variable is totally superfluous. In this case, the covariance matrix is not positive definite. Its true dimension, or *rank,* is 1, which means that it has only one meaningful risk factor.

These conditions may seem academic, but unfortunately, they soon become very real with simulations based on a large number of factors. The RiskMetrics covariance matrix, for instance, is routinely nonpositive definite owing to the large number of assets. These problems can arise for a number of reasons. Perhaps this is simply due to the large number of correlations. With $N = 450$, for instance, we have about 100,000 correlations with rounding errors. This also could happen when the effective number of observations T is less than the number of factors N. One drawback of time-varying models of variances is that they put less weight on older observations, thereby reducing the effective sample size. Or the correlations may have been measured over different periods, which may produce inconsistent correlations.[9] Another reason would be that the series are naturally highly correlated (such as the 9-year zero-coupon bond with the adjoining maturities) or that some series were constructed as a linear combination of others (such as a currency basket).

For simulations, this may be a blessing in disguise because fewer number of variables are sufficient. In Chapter 8 we gave the example of 11 bonds for which the covariance matrix could be reduced without much loss of information to two, or perhaps three, principal components. Thus the problem can be solved using a matrix of smaller dimensions, which speeds up the computation considerably. This illustrates that the design of simulation experiments, including the number of risk factors, is critical. As we have seen, however, the choice of the number of risk factors should be related to the trading strategy.

12.5 DETERMINISTIC SIMULATION

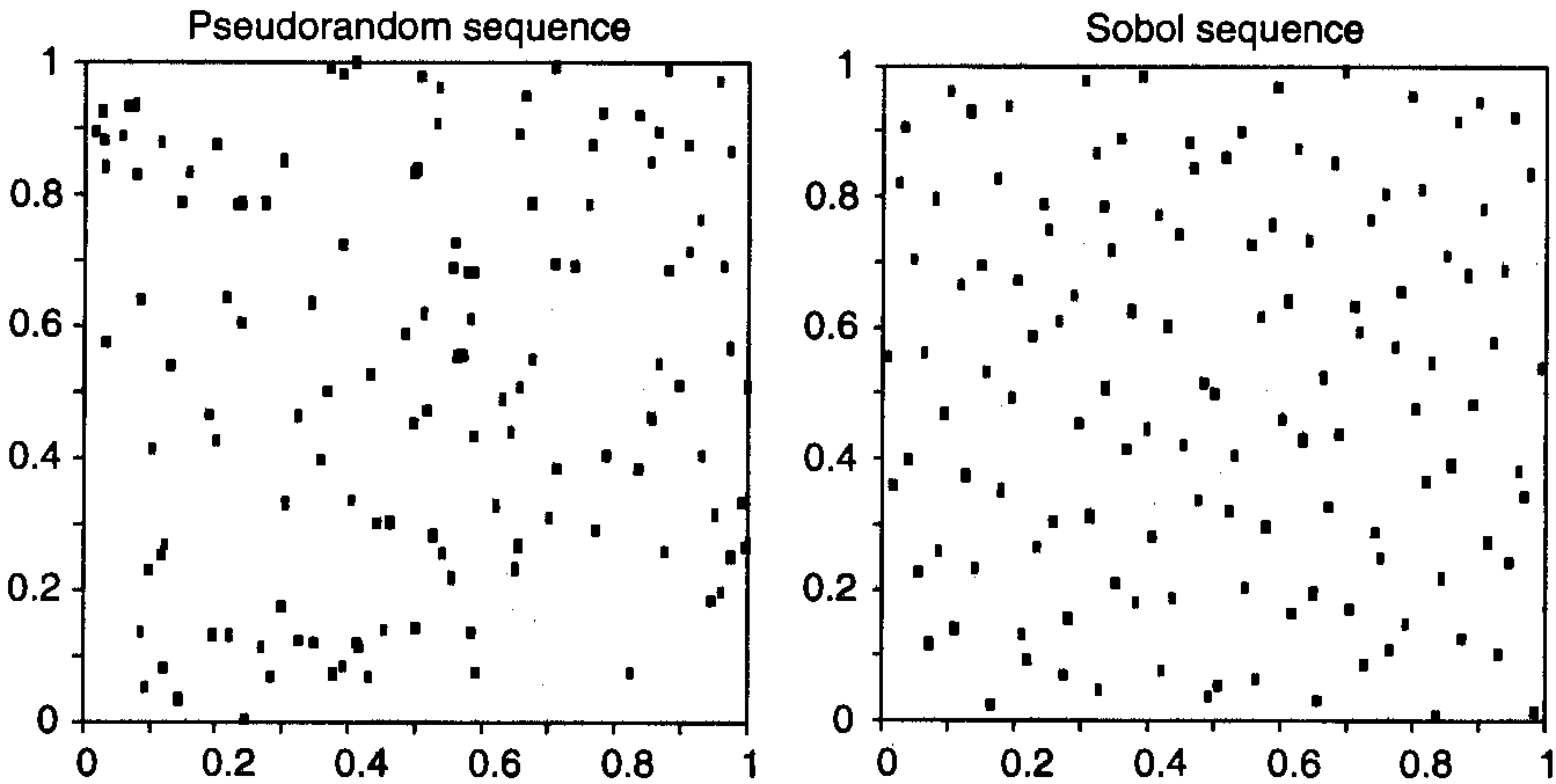
Monte Carlo simulation methods generate independent, pseudorandom points that attempt to "fill" an N-dimensional space, where N is the number of risk factors. The sequence of points does not have to be chosen randomly, however.

Indeed, it is possible to use a *deterministic* scheme that is constructed to provide a more consistent fill to the N-space. The choice must account for the sample size, dimensionality of the problem, and possibly the shape of the function being integrated. These deterministic schemes are sometimes called *quasi–Monte Carlo* (QMC), although this is a misnomer because there is nothing random about them. The numbers are not independent but rather are constructed as an ordered sequence of points.

[9] As an example, consider three assets. We have 1 year of daily data for assets A and B, which have a high correlation, say, 0.9. For asset C, we only have 1 month of data. If, over the shorter period, asset C has a high observed correlation with A and a low enough correlation with B, the correlation matrix will be inconsistent. For instance, if the correlation between C and A is 0.9, the lowest possible correlation between C and B is 0.62.

FIGURE 12-4

Comparison of distributions.



To illustrate, Figure 12-4 compares a distribution for two variables only (after all, this is the number of dimensions of a page). The figure shows, on the left, pseudorandom points and, on the right, a deterministic, *low-discrepancy* sequence obtained from a so-called Sobol procedure.[10]

The left graph shows that the points often "clump" in some regions and leave out large areas. These clumps are a waste because they do not contribute more information. The right panel, in contrast, has more uniform coverage. Instead of drawing independent samples, the deterministic scheme systematically fills the space left by the previous numbers in the series.

Quasirandom methods have the desirable property that the standard error shrinks at a faster rate, proportional to close to $1/K$ rather than $1/\sqrt{K}$ for standard simulations. Indeed, a number of authors have shown that deterministic methods provide a noticeable improvement in speed.[11] Papageorgiou and Paskov (1999) compare the computation of VAR for a portfolio exposed to 34 risk factors using 1000 points. They find that the deterministic sequence can be 10 times more accurate than the Monte Carlo method.

[10] This algorithm is described in Press et al. (1992).

[11] See, for example, Boyle et al. (1997) for call options and Paskov and Traub (1995) for mortgage securities.

One drawback of these methods is that since the draws are not independent, accuracy cannot be assessed easily. For the Monte Carlo method, in contrast, we can construct confidence bands around the estimates. Another issue is that for high-dimensionality problems, some QMC sequences tend to cycle, which leads to decreases in performance. Overall, however, suitably selected QMC methods can provide substantial accelerations in the computations.

12.6 CHOOSING THE MODEL

Simulation methods are most prone to model risk. If the stochastic process chosen for the price is unrealistic, so will be the estimate of VAR. This is why the choice of the underlying process is particularly important.

For example, the geometric brownian motion model in Equation (12.1) adequately describes the behavior of some financial variables, such as stock prices and exchange rates, but certainly not that of fixed-income securities. In the brownian motion models, shocks to the price are never reversed, and prices move as a random walk. This cannot represent the price process for default-free bond prices, which must converge to their face value at expiration.

Another approach is to model the dynamics of interest rates as

$$dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} dz_t \qquad (12.10)$$

This class of model includes the Vasicek (1977) model when $\gamma = 0$; changes in yields then are normally distributed, which is particularly convenient because this leads to many closed-form solutions. With $\gamma = 0.5$, this is also the Cox, Ingersoll, and Ross (1985) model of the term structure (CIR). With $\gamma = 1$, the model is lognormal.[12]

This process is important because it provides a simple description of the stochastic nature of interest rates that is consistent with the empirical observation that interest rates tend to be mean-reverting. Here, the parameter $\kappa < 1$ defines the speed of mean reversion toward the long-run value θ. Situations where current interest rates are high, such as $r_t > \theta$, imply a negative drift $\kappa(\theta - r_t)$ until rates revert to θ. Conversely, low current rates are associated with positive expected drift. Also note that with $\gamma = 0.5$, the variance of this process is proportional to the level of

[12] Bliss and Smith (1998) show that $\gamma = 0.5$ provides a good fit to U.S. short-term interest rates.

interest rates; as the interest rate moves toward 0, the variance decreases, so r can never fall below 0. If the horizon is short, however, the trend or mean reversion term will not be important.

Equation (12.10) describes a one-factor model of interest rates that is driven by movements in short-term rates dr_t. In this model, movements in longer-term rates are perfectly correlated with movements in this short-term rate through dz. Therefore, the correlation matrix of zero-coupon bonds consists of ones only. This may be useful to describe the risks of some simple portfolios but certainly not for the leveraged fixed-income portfolios of financial institutions.

For more precision, additional factors can be added. Brennan and Schwartz (1979), for example, proposed a two-factor model with a short and long interest rate modeled as

$$dr_t = \kappa_1(\theta_1 - r_t)dt + \sigma_1 dz_{1t} \tag{12.11}$$

$$dl_t = \kappa_2(\theta_2 - l_t)dt + \sigma_2 dz_{2t} \tag{12.12}$$

where l is the long rate and the errors have some correlation. Generalizing, one could use the full covariance matrix along some 14 points on the yield curve, as provided by RiskMetrics. In theory, *finer granularity* should result in better risk measures, albeit at the expense of computational time. In all these cases, the Monte Carlo experiment consists of first simulating movements in the driving risk factors and then using the simulated term structure to price the securities at the target date.

Here is where risk management differs from valuation methods. For short horizons (say, 1 day to 1 month), we could assume that changes in yields are jointly normally distributed. This assumption may be quite sufficient for risk management purposes. Admittedly, it would produce inconsistencies over long horizons (say, beyond a year) because yields could drift in different directions, creating term structures that look unrealistic.[13]

With longer horizons, the drift term in Equation (12.11), for example, becomes increasingly important. To ensure that the two rates cannot move too far away from each other, one could incorporate into the drift of the short rate an *error-correction term* that pushes the short rate down when it is higher than the long rate. For instance, one could set

[13] This explains why the Black model is used often to price short-term options on long-term bonds. Although theoretically inconsistent, it produces good results because the maturity of the option is so short relative to that of the underlying.

$$dr_t = \kappa_1[\theta_1 - (r_t - l_t)]dt + \sigma_1 dz_{1t} \tag{12.13}$$

Indeed, much work has been devoted to the analysis of time series that are *cointegrated*, that is, that share a common random component.[14] These error-correction mechanisms can be applied to larger-scale problems, thus making sure that our 14 yields move in a realistic fashion.

But again, over short horizons, the modeling of expected returns is not too important. This also applies to the choice of term structure models, equilibrium models versus arbitrage models. *Equilibrium models* postulate a stochastic process for some risk factors, which generates a term structure. This term structure, however, will not fit exactly the current term structure, which is not satisfactory for fixed-income option traders. They argue that if the model does not even fit current bond prices, it cannot possibly be useful to describe options. This is why *arbitrage models* take today's term structure as an input (instead of output for the equilibrium models) and fit the stochastic process accordingly.

For instance, a one-factor no-arbitrage model is

$$dr_t = \theta(t)dt + \sigma dz_t \tag{12.14}$$

where the function $\theta(t)$ is chosen so that today's bond prices fit the current term structure. This approach has been extended to two-factor Heath-Jarrow-Morton (1992) models, but their estimation is computer-intensive and has been described "at the very boundaries of feasibility."[15] These *arbitrage* models are less useful for risk management, however.

For risk management purposes, what matters is to capture the richness in movements in the term structure, not necessarily to price today's instruments to the last decimal point. Thus the "art" of risk management lies in deciding what elements of the model are important.

12.7 CONCLUSIONS

Simulation methods are now used widely for risk management purposes. Interestingly, these methods can be traced back to the valuation of complex options, except that there is no discounting or risk-neutrality assumption. Thus the investment in intellectual and systems development for derivatives

[14] Much of the groundbreaking work in cointegration was done by Engle and Granger while at the University of California at San Diego. See Engle and Granger (1991), a good review book.

[15] See Rebonato (1996).

TABLE 12-4

Comparison of VAR Methods

	Valuation Method	
Risk-Factor Distribution	**Local Valuation**	**Full Valuation**
Variance-covariance	Delta-normal	
Historical simulation		Historical path
Deterministic simulation		Full simulation
Monte Carlo simulation	Delta-gamma-MC	Full Monte Carlo

trading can be used readily for computing VAR. No doubt this is why officials at the Fed have stated that derivatives "have had favorable spillover effects on institutions' abilities to manage their total portfolios."

Simulation methods are quite flexible. They can either postulate a stochastic process or resample from historical data. They allow full valuation on the target date. On the downside, they are more prone to model risk owing to the need to prespecify the distribution and are much slower and less transparent than analytical methods. In addition, simulation methods create sampling variation in the measurement of VAR. Greater precision comes at the expense of vastly increasing the number of replications, which slows the process down.

VAR methods are listed in Table 12-4 in order of increasing time requirement. At one extreme is the Monte Carlo method, which requires the most computing time. For the same accuracy, deterministic simulations are faster because they create more systematic coverage of the risk factors. Next is the historical simulation method, which uses recent history in a limited number of simulations. At other extreme is the delta-normal method, which requires no simulation and is very fast. With the ever-decreasing cost of computing power and advances in scientific methods, however, we should expect greater use of simulation methods.

QUESTIONS

1. What is the main assumption for the risk factors underlying the Monte Carlo simulation method?
2. What is the main assumption for the risk factors underlying the historical simulation method?

3. Explain why numerical integration is plagued by the curse of dimensionality and why this is avoided by the Monte Carlo simulation method.
4. Define K as the number of Monte Carlo replications. At what rate does the standard error of estimates decrease?
5. What are the major drawbacks of the Monte Carlo simulation method?
6. Consider an operating system that has a random-number generator with a short cycle, that is, that repeats the same sequence of numbers after a few thousand iterations. Will this lead to inaccuracy in the calculation of VAR? Why?
7. Explain how the inverse transform method could generate draws from a student t distribution.
8. What is an advantage of the bootstrap approach compared with a Monte Carlo simulation based on the normal distribution?
9. If the movements in the risk factors have positive autocorrelation from one day to the next, can we bootstrap on the changes in the risk factors?
10. To compute VAR using a simulation method, which two statistics are required?
11. Can Monte Carlo simulation be adapted to changing volatility?
12. Explain why pricing methods use risk-neutral distributions. Does risk measurement need risk-neutral or physical distributions?
13. A Monte Carlo simulation creates a 99 percent VAR estimate of \$10 million with a standard error of \$4 million using 1000 replications. How many replications are needed to shrink this standard error to less than \$1 million?
14. The relative error in the previous question was 4/10. Would you expect this ratio to be higher or lower for a 95 percent VAR?
15. How many years of daily data do we need to estimate a 99 percent VAR with a precision of 1 percent or better? Would we need more/fewer years for distributions with negative skewness, and why?
16. Explain how stratified sampling could generate more precise estimates of VAR.
17. A risk manager needs to generate two variables with a correlation of 0.6. Explain how this could be done starting from independent variables. Verify that the final variables have unit volatility.
18. Assume now that the correlation between the two variables is 1. What does this imply in terms of independent risk factors for the simulation?

19. Are sequences of variables in the quasi–Monte Carlo methods independent?
20. Is the geometric brownian motion model a good description of the behavior of fixed-income securities?
21. Explain how equilibrium and no-arbitrage models use the current term structure.

CHAPTER 13

Liquidity Risk

> LTCM then faced severe market liquidity problems when its investments began losing value and the fund attempted to unwind some of its positions.
>
> —*President's Working Group on Financial Markets, 1999*

Traditional value-at-risk (VAR) models assume that the portfolio is "frozen" over the horizon and that market prices represent achievable transaction prices. This marking-to-market approach is adequate to quantify and control risk for an ongoing portfolio but may be more questionable if VAR is supposed to represent the worst loss over a liquidation period.

The question is how VAR can be adapted to deal with liquidity considerations. As we saw in Chapter 1, liquidity risk can be grouped into *funding liquidity risk* and *asset liquidity risk.* Funding liquidity risk arises when financing cannot be maintained owing to creditor or investor demands. The resulting need for cash may require selling assets. Asset liquidity risk arises when a forced liquidation of assets creates unfavorable price movements. Thus liquidity considerations should be viewed in the context of both the assets and the liabilities of the financial institution.

This chapter discusses recent developments that adapt traditional VAR measures to liquidity considerations. Section 13.1 first provides a general introduction to asset and funding liquidity risk.

Next, Section 13.2 attempts to incorporate asset liquidity risk into VAR measures. Immediate liquidation can create losses owing to market impact, which is a drop in the liquidation value relative to mark-to-market prices. Liquidation, however, can take place over many days and should be done so as to balance transactions costs and price risk. Taking both costs and risk into account leads to a measure of liquidity-adjusted VAR (LVAR). Often, however, liquidity is factored into the valuation of positions by decreasing their value by a reserve.

Section 13.3 then discusses measures of funding liquidity risk proposed by the Counterparty Risk Management Policy Group (CRMPG). Even though an institution can have zero traditional VAR, different swap credit terms can generate different cash requirements. The cash liquidity measure is an extension of VAR.

Next, Section 13.4 is devoted to an analysis of the Long-Term Capital Management (LTCM) debacle. LTCM failed because of its lack of diversification combined with its asset and funding liquidity risk, which were due to its sheer size.

Finally, Section 13.5 provides some concluding comments about liquidity risk. Liquidity problems have proved to be crucial in the failure of many financial institutions. Liquidity risk can be factored formally into hybrid VAR measures but only using price impact functions derived from normal market conditions. During episodes of systemic risk, however, liquidity evaporates, invalidating much of this analysis. Thus liquidity risk probably is the weakest spot of market risk management systems.

13.1 DEFINING LIQUIDITY RISK

Table 13-1 displays sources of liquidity risk for a financial institution. Liquidity risk emanates from the liability side, when creditors or investors demand their money back. This usually happens after the institution has incurred or is thought to have incurred losses that could threaten its solvency. The need for cash creates problems on the asset side when the forced liquidation of assets causes transactions losses.

Understanding liquidity risk requires knowledge of several different fields, including *market microstructure*, which is the study of market-

TABLE 13-1

Sources of Liquidity Risk

Assets	**Liabilities**
Size of position	Funding
Price impact for unit trade	Mark to market, haircuts
	Equity
	Investor redemptions

clearing mechanisms; *optimal trade execution*, which is the design of strategies to minimize trading costs or to meet some other objective function; and *asset-liability management*, which attempts to match the values of assets and liabilities on balance sheets.

13.1.1 Asset Liquidity Risk

Asset liquidity risk, sometimes called *market/product liquidity risk*, is the risk that the liquidation value of assets may differ significantly from their current mark-to-market values. It is a function of the price impact of trades and the size of the positions.

Asset liquidity can be measured by a price-quantity function. This is also known as the *market-impact* effect. Highly liquid assets, such as major currencies or Treasury bonds, are characterized by *deep markets*, where positions can be offset with very little price impact. *Thin* markets, such as exotic over-the-counter (OTC) derivatives contracts or some emerging market equities, are those where any transaction can quickly affect prices.

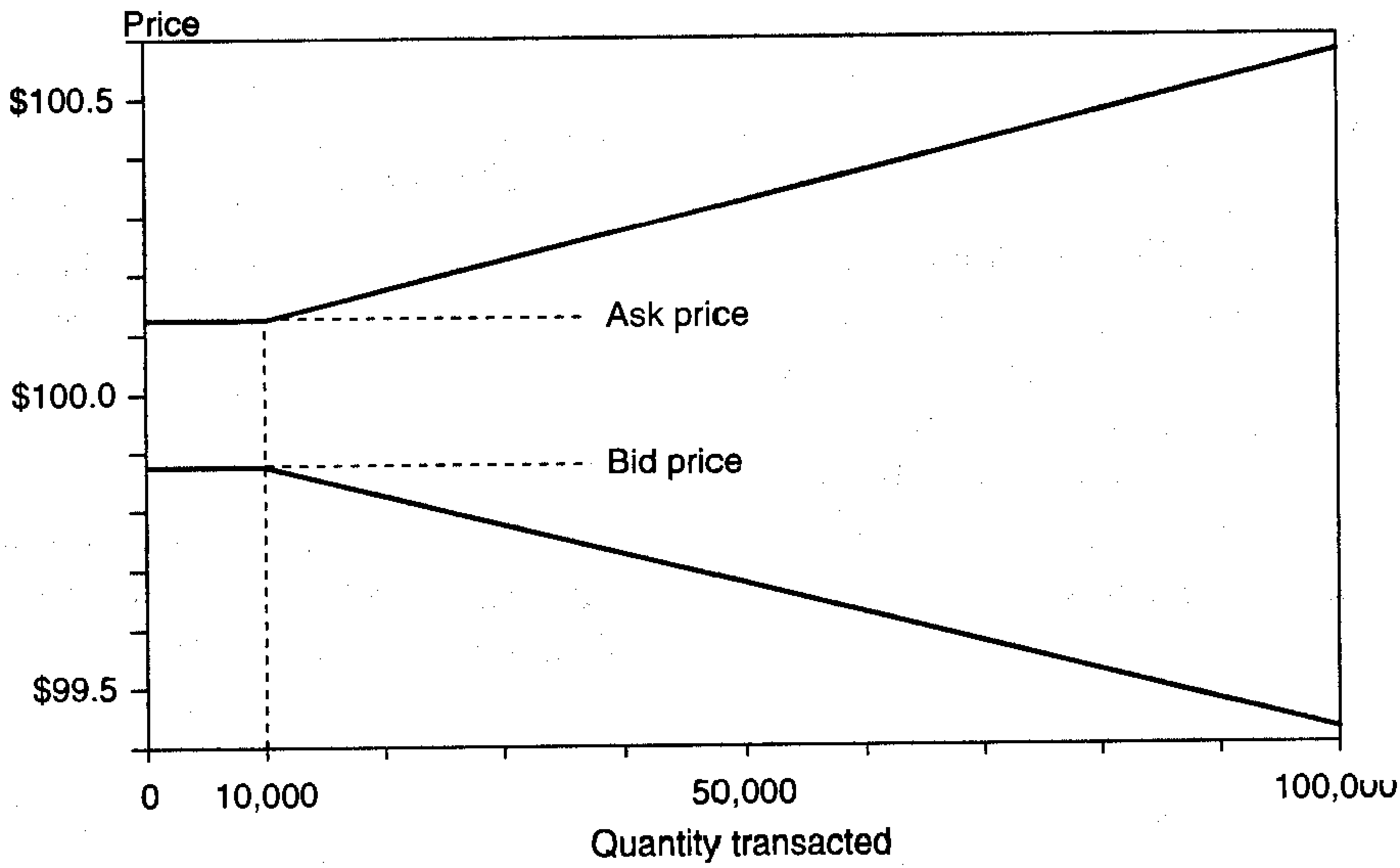
This price function is illustrated in Figure 13-1.[1] The starting point is the current *midprice*, which is the average of the bid and ask quotes and can be used to mark the portfolio to market. Here, the *bid-ask spread* is \$0.25. Markets with low spreads are said to exhibit *tightness*. In Figure 13-1, this is valid up to some limit, say, 10,000 shares. This is sometimes called the *normal market size*, or *depth*. Relative to mid-market values, the cost of trading is half the spread. This component of trading cost is sometimes called *exogenous* because it does not depend on quantities transacted, as long as these quantities are below the normal market size.

For quantities beyond this point, however, the sale price is a decreasing function of the quantity, reflecting the price pressure required to clear the market. The converse is true for the purchase price. In practice, the position is compared with some metric such as the median daily trading volume. For a widely traded stock such as IBM, for instance, selling 4 percent of the daily trading volume incurs a cost of about 60 basis points. Studies of market microstructure provide empirical evidence on trading costs.

[1] In what follows, we ignore the fixed component of trading costs, that is, commissions and taxes.

FIGURE 13-1

Price-quantity function.



This relationship is assumed to be linear, although it could take another shape. The slope of the line measures the *market impact*. This varies across assets and, possibly, across time for a given asset. In this example, selling 100,000 shares over 1 day would push the price down from a midmarket value of \$100 to about \$99.4. Thus, selling this position would incur a liquidation cost of (\$100−\$99.4)/\$100, or 60 basis points.

This demonstrates that liquidity depends on both the price-impact function and the size of the position. In this example, if the position is below 10,000 shares, then market liquidity is not a major issue. In contrast, if the institution holds a number of shares worth several days of normal trading volume, liquidity should be of primary concern.

In addition to varying across assets, liquidity is also a function of prevailing market conditions. This is more worrying because markets seem to go through regular bouts of liquidity crises. Most notably, liquidity in bond markets dried up during the summer of 1998 as uncertainty about defaults led to a "flight to quality," that is, increases in the prices of Treasuries relative to those of other bonds. A similar experience occurred during the 1994 bond market debacle, at which time it became quite difficult to deal in Eurobonds or mortgage-backed securities.

Traditionally, asset liquidity risk has been controlled through position limits. The goal of *position limits* is to limit the exposure to a single instrument, even if it provides diversification of market risk, in order to avoid a large market impact in case of forced liquidation.

13.1.2 Funding Liquidity Risk

Cash-flow/funding liquidity risk refers to the inability to meet payment obligations to creditors or investors. This can force unwanted liquidation of the portfolio.

Funding risk arises from the liability side of the balance sheet. Most financial institutions are *leveraged*. Often, this involves posting some collateral (assets) in exchange for cash from a broker. Normally, brokers require collateral that is worth slightly more than the cash loaned, by an amount known as a *haircut*, designed to provide a buffer against decreases in the collateral value. The value of the collateral, however, is constantly *marked to market* by the broker. If this value falls, the broker will require some additional payment, called *variation margin*, to keep the total amount held above the loan value. If the institution does not have enough cash on hand, it will be forced to liquidate some of its other assets.

Brokers also reserve the right of *changes in collateral requirements*, which can create additional cash flow risk. For example, brokers can increase the haircut when markets are more volatile, creating extra demands on cash. Similarly, organized exchanges can change their required margins at will.

Finally, cash-flow liquidity risk also arises owing to *mismatches in the timing of payments*. Even if an institution is perfectly matched in terms of market risk, it may be forced to make a payment on a position without having yet received an offsetting payment on a hedge. Section 13.3 gives examples of such mismatches.

The first line of defense against funding liquidity risk is *cash*. Another may be a *line of credit*, which is a loan arrangement with a bank allowing the customer to borrow up to a prespecified amount.

The institution may be able to meet margin calls by raising funds from another source, such as new debt or a new equity issue. In practice, it may be difficult to raise new funds precisely when the institution is faring badly and needing them most.

Conversely, the institution must evaluate the likelihood of redemptions, or cash requests from debt holders or equity holders. This is most likely to occur when the institution appears most vulnerable, thereby transforming what could be a minor problem into a crisis. It is also important to avoid debt covenants or options that contain "triggers" that would force early redemption of the borrowed funds. Such credit triggers accelerated the fall of Enron, as shown in Box 13-1.

BOX 13-1

ENRON'S CREDIT TRIGGERS: THE BAD AND STUPID

Credit triggers are clauses in financial contracts that allow creditors to demand immediate payments if the credit rating of the borrower falls below some predetermined level.

Enron is now widely viewed as a massive case of accounting fraud. Credit triggers, however, played a role in Enron's demise. Enron was rated investment-grade until November 28, 2001, when a proposed takeover by Dynergy fell through. On that day, Standard & Poor's downgraded Enron to speculative-grade, triggering the immediate repayment of almost $4 billion in debt. Unable to pay, Enron filed for bankruptcy on December 2, 2001.

The real cause of Enron's failure was its poor performance in many business lines, which was hidden through creative off-balance-sheet financing. As early as 1999, Vince Kaminsky, Enron's risk manager, had railed against these arrangements, which he said had gone from merely "stupid" to fraudulent. His comments, unfortunately, were ignored by top management.

Ostensibly, credit triggers are designed to lower the cost of capital for the issuing company. Because this is an option granted to debt holders, they should be willing to accept a lower interest rate than otherwise. Superficially, such clauses look beneficial because lenders can *put* the obligation back to the borrower.

In practice, however, there have been many cases where credit triggers offered no protection to creditors because they precipitated a default. In such situations, from the viewpoint of borrowers, the cost savings have been swamped by the problems caused by credit triggers.

Credit-rating agencies call these credit triggers "problematic" and "troubling." As a result, they now examine much more closely the potential effects of credit triggers and take them into account when setting ratings.*

* See Moody's (2001).

Thus liquidity considerations should be viewed in the context of both asset and liabilities. Consider, for instance, *hedge funds*, some of which invest in illiquid assets such as distressed debt. To minimize liquidity risk, such funds impose a longer *lockup period*, or minimum time for investors to keep their funds, and a longer *redemption notice period* for withdrawing funds.

As explained in Chapter 3, *commercial banks* are by their nature susceptible to liquidity risks. They are funded by short-term deposits but can invest in illiquid real estate loans. This setup is fraught with liquidity risk and explains the rationale for deposit insurance, which eliminates the incentives for bank runs.

13.2 ASSESSING ASSET LIQUIDITY RISK

Trading returns are measured typically from midmarket prices. This may be adequate for measuring daily profit and loss (P&L) but may not represent the actual fall in value if a large portfolio were to be liquidated. The question is how to assess potential losses under such conditions.[2] In turn, this can give insights into how to manage this risk.

Traditional adjustments are done on an ad hoc basis. Liquidity risk can be loosely factored into VAR measures by ensuring that the *horizon* is at least greater than an orderly liquidation period. Generally, the same horizon is applied to all asset classes, even though some may be more liquid than others.

Sometimes, longer liquidation periods for some assets are taken into account by artificially increasing the volatility. For instance, one could mix a large position in the dollar/yen with another one in the dollar/Polish zloty, both of which have an annual volatility of 10 percent, by artificially increasing the volatility of the second foreign currency in the VAR computations.

13.2.1 Effect of Bid-Ask Spreads

More formally, one can focus on the various components of liquidation costs. The first and most easily measurable is the quoted bid-ask spread, defined in relative terms, that is,

[2] Note that the section is titled "Assessing Asset Liquidity Risk" instead of "Measuring Asset Liquidity Risk." This is to reflect the fact that assessment is less precise than measurement.

$$S = \frac{[P(\text{ask}) - P(\text{bid})]}{P(\text{mid})} \tag{13.1}$$

Table 13-2 provides typical spreads. We see that spreads vary from a low of about 0.05 percent for major currencies, large U.S. stocks, and on-the-run Treasuries to much higher values when dealing with less liquid currencies, stocks, and bonds. Treasury bills are in a class of their own, with extremely low spreads. These spreads are indicative only because they depend on market conditions. Also, market makers may be willing to trade within the spread.

At this point, it is useful to review briefly the drivers of these spreads. According to market microstructure theory, spreads reflect three different types of costs:

- *Order-processing costs* cover the cost of providing liquidity services and reflect the cost of trading, the volume of transaction, the state of technology, and competition. With fixed operating costs, these order-processing costs should decrease with transaction volumes.

TABLE 13-2

Typical Spreads and Volatility

		Volatility (%)	
Asset	**Spread (%) (Bid-Ask)**	**Daily**	**Annual**
Currencies			
Major (euro, yen, . . .)	0.02–0.10	0.3–1.0	5–15
Emerging (floating)	0.10–1.00	0.3–1.9	5–30
Bonds			
On-the-run Treasuries	0.03	0.0–0.7	0–11
Off-the-run Treasuries	0.06–0.20	0.0–0.7	0–11
Corporates	0.10–1.00	0.0–0.7	0–11
Treasury bills	0.003–0.02	0.0–0.1	0–1
Stocks			
U.S.	0.05–5.00	1.3–3.8	20–60
Average, NYSE	0.20	1.0	15
Average, all countries	0.40	1.0–1.9	15–30

Note: Author's calculations. Cost of trades excludes broker commissions and fees. See also *Institutional Investor* (November 1999).

- *Asymmetric-information costs* reflect the fact that some orders may come from informed traders, at the expense of market makers who can somewhat protect themselves by increasing the spread.
- *Inventory carrying costs* are due to the cost of maintaining open positions, which increase with higher price volatility, higher interest-rate carrying costs, and lower trading activity or turnover.

If the spread were fixed, one simply could construct a liquidity-adjusted VAR from the traditional VAR by adding a term, that is,

$$\text{LVAR} = \text{VAR} + L_1 = (W\alpha\sigma) + 1/2(WS) \qquad (13.2)$$

where W is the initial wealth, or portfolio value. For instance, if we have \$1 million invested in a typical stock with a daily volatility of $\sigma = 1$ percent and spread of $S = 0.20$ percent, the 1-day LVAR at the 95 percent confidence level would be

$$(\$1{,}000{,}000 \times 1.645 \times 0.01) + 1/2\,(\$1{,}000{,}000 \times 0.0020)$$
$$= \$16{,}450 + \$1000 = \$17{,}450$$

Here, the correction factor is relatively small, accounting for 7 percent of the total.

This adjustment can be repeated for all assets in the portfolio, leading to a series of add-ons, $1/2\,\Sigma_i|W_i|S_i$. This sequence of positive terms increases linearly with the number of assets, whereas the usual VAR benefits from diversification effects. Thus the relative importance of the correction factor will be greater for large portfolios.

A slightly more general approach is proposed by Bangia et al. (1999), who consider the uncertainty in the spread. They characterize the distribution by its mean $\bar{S}$ and standard deviation σ_S. The adjustment considers the worst increase in the spread at some confidence level, that is,

$$\text{LVAR} = \text{VAR} + L_2 = (W\alpha\sigma) + \tfrac{1}{2}\,[W(\bar{S} + \alpha'\sigma_s)] \qquad (13.3)$$

This assumes that the worst market loss and increase in spread will occur simultaneously. In general, we observe a positive correlation between volatility and spreads.

At the portfolio level, one theoretically could take into account correlations between spreads. In practice, summing the individual worst spreads provides a conservative measure of the portfolio worst spread.

Typically, σ_s is about half the size of the average spread; for example, $\sigma_s = 0.02$ percent against $\overline{S} = 0.04$ percent for the dollar/euro exchange rate. Relative to a volatility of about 1.0 percent per day, these adjustments are small. Thus transactions costs based on spreads are not very important relative to usual VAR measures.

13.2.2 Incorporating Liquidity in Valuation

If the position is to be sold, the second term in Equation (13.2) represents a certain loss, unlike the volatility term. Assuming that the portfolio is valued using midmarket prices, it represents the loss owing to the liquidation.

Another approach to liquidity is to mark the portfolio to the appropriate bid prices (for long positions) or ask prices (for short positions). In practice, financial institutions generally mark their cash positions to the conservative bid-offer basis.[3] VAR then can be viewed as the worst loss from this value.

Further, financial institutions often apply *reserves*, which are pricing changes in the valuation away from fair value to account for such effects as illiquidity and model risk. Firms deduct this reserve from the fair value of positions to account for the extra time and cost required to close out the position. The reserve amount is often based on judgments about the liquidity of a market. In such cases, there is no need to take liquidity risk into account in VAR because it is already factored into the valuation of positions.

13.2.3 Effect of Price Impact

Although this approach has the merit of considering some transactions costs, it is not totally satisfactory. It only looks at the bid-ask spread component of these costs, which may be appropriate for a small portfolio but certainly not when liquidation can affect market prices. The market-impact factor should be taken into account.

To simplify, let us assume a linear price-quantity function and ignore the spread. For a sale, the price per share is

$$P(q) = P_0(1 - kq) \tag{13.4}$$

[3] See the Basel Committee (2005a) survey of financial institutions. Derivatives, however, generally are marked to midmarket values.

Assume that $P_0 = \$100$ and $k = 0.5 \times 10^{-7}$. Say that we start with a position of $q = 1$ million shares of the stock. If we liquidate all at once, the price drop will be $P_0 kq = \$100 \times (0.5 \times 10^{-7}) \times 1{,}000{,}000 = \$5$ per share, leading to a total price impact of \$5 million. In contrast, we could decide to work the order through at a constant rate of 200,000 shares over $n = 5$ days. In the absence of other price movements, the daily price drop will be \$1 per share, leading to a total price impact of \$1 million, much less than before.

Immediate liquidation creates the costs:

$$C_1(W) = q \times [P_0 - P(q)] = q \times (P_0 - P_0 + P_0 kq) = kq^2 P_0 \quad (13.5)$$

Uniform liquidation creates the costs:

$$C_2(W) = q \times [P_0 - P(q/n)] = q \times (P_0 - P_0 + P_0 kq/n) = k(q^2/n) P_0 \quad (13.6)$$

Because uniform liquidation spreads the price impact over many days, it leads to lower trading costs.

The drawback of liquidating more slowly, however, is that the portfolio remains exposed to price risks over a longer period. The position profiles are compared in Figure 13-2. Under the immediate sale, the

FIGURE 13-2

Profile of execution strategies.