weights decay fairly quickly, dropping below 0.00012 for data more than 100 days old. Thus the number of *effective* observations is rather small.

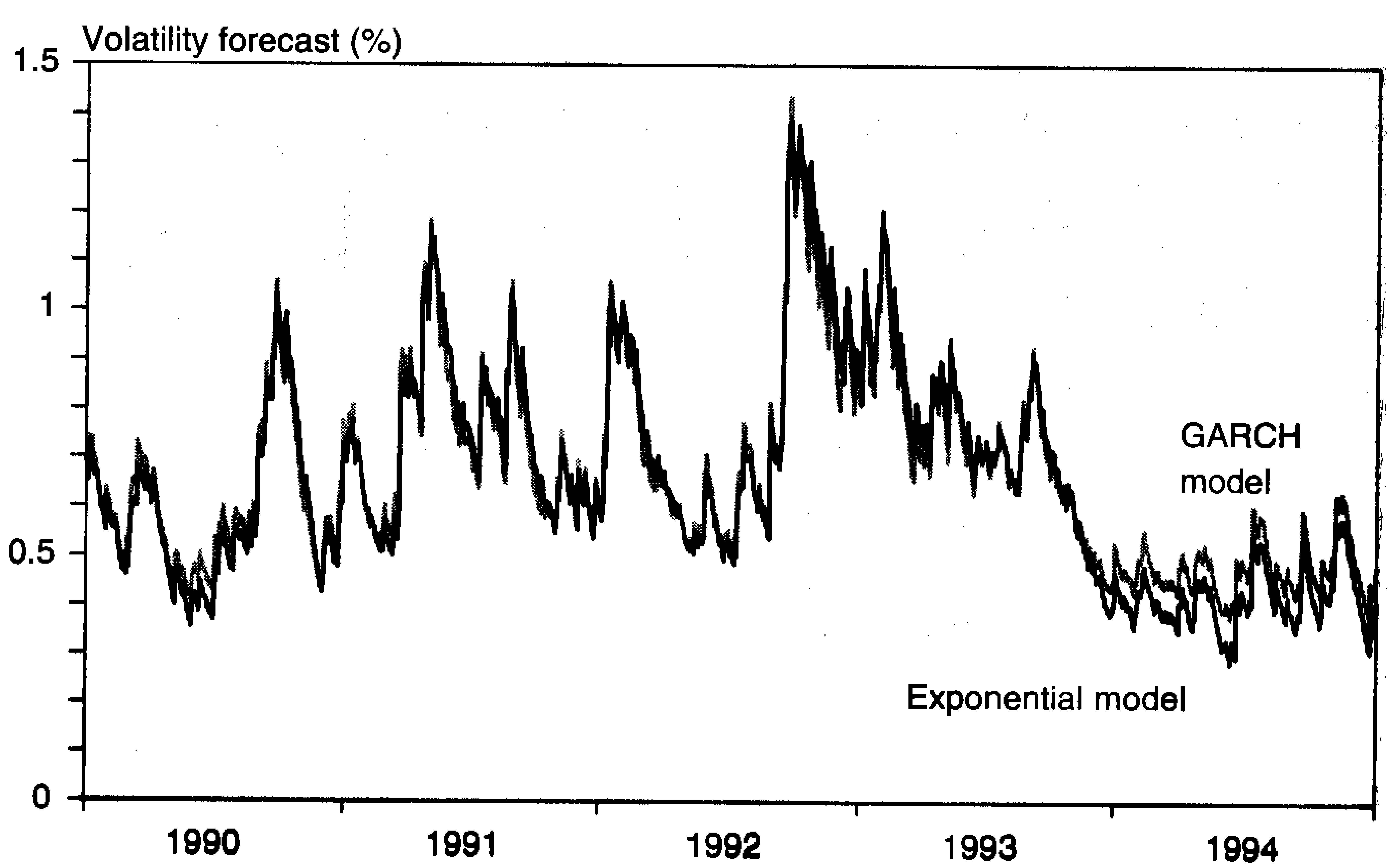
This model is a special case of the GARCH process where α_0 is set to 0 and α_1 and β sum to unity. The model therefore has persistence of 1. It is called *integrated GARCH* (IGARCH). As shown in Figure 9-8, the 1-day forecasts are nearly identical to those obtained with the GARCH model in Figure 9-4. The longer-period forecasts, however, are markedly different because the EWMA process does not revert to the mean.

The exponential model is particularly easy to implement because it relies on one parameter only. Thus it is more robust to estimation error than other models. In addition, as was the case for the GARCH model, the estimator is *recursive*; the forecast is based on the previous forecast and the latest innovation. The whole history is summarized by one number, h_{t-1}. This is in contrast to the moving average, for instance, where the last M returns must be used to construct the forecast.

The only parameter in this model is the decay factor λ. In theory, this could be found from maximizing the likelihood function. Operationally, this would be a daunting task to perform every day for hundreds of time series. An optimization has other shortcomings. The decay factor may vary

FIGURE 9-8

Exponential volatility forecast.



not only across series but also over time, thus losing consistency over different periods. In addition, different values of λ create incompatibilities across the covariance terms and may lead to unreasonable values for correlations, as we shall see later. In practice, RiskMetrics only uses one decay factor for all series, which is set at 0.94 for daily data.

RiskMetrics also provides risk forecasts over monthly horizons, defined as 25 trading days. In theory, the 1-day exponential model should be used to extrapolate volatility over the next day, then the next, and so on until the twenty-fifth day ahead, as was done for the GARCH model earlier. Herein lies the rub.

The persistence parameter for the exponential model $(\alpha_1 + \beta)$ is unity. Thus the model allows no mean reversion, and the monthly volatility should be the same as the daily volatility. In practice, however, we do observe mean reversion in monthly risk forecasts.

This is why RiskMetrics takes a different approach. The estimator uses the same form as Equation (9.9), redefining r_{t-1} as the 25-day moving variance estimator, that is,

$$h'_t = \lambda h'_{t-1} + (1-\lambda)s^2_{t-1}, \quad s^2_{t-1} = \sum_{k=1}^{25} r^2_{t-k} \tag{9.11}$$

In practice, this creates strange "ghost" features in the pattern of monthly variance forecast.

After experimenting with the data, J.P. Morgan chose $\lambda = 0.97$ as the optimal decay factor. Therefore, the daily and monthly models are inconsistent with each other. However, they are both easy to use, they approximate the behavior of actual data quite well, and they are robust to misspecification.

9.3 MODELING CORRELATIONS

Correlation is of paramount importance for portfolio risk, even more so than individual variances. To illustrate the estimation of correlation, we pick two series: the dollar/British pound exchange rate and the dollar/Deutsche mark rate.

Over the 1990–1994 period, the average daily correlation coefficient was 0.7732. We should expect, however, some variation in the correlation coefficient because this time period covers fixed and floating exchange-rates regimes. On October 8, 1990, the pound became pegged to the mark within the European Monetary System (EMS). This lasted until the turmoil of

September 1992, during which sterling left the EMS and again floated against the mark.

As in the case of variance estimation, various methods can be used to capture time variation in correlation: moving average, GARCH, and exponential. Correlations can be derived from *multivariate* GARCH models.

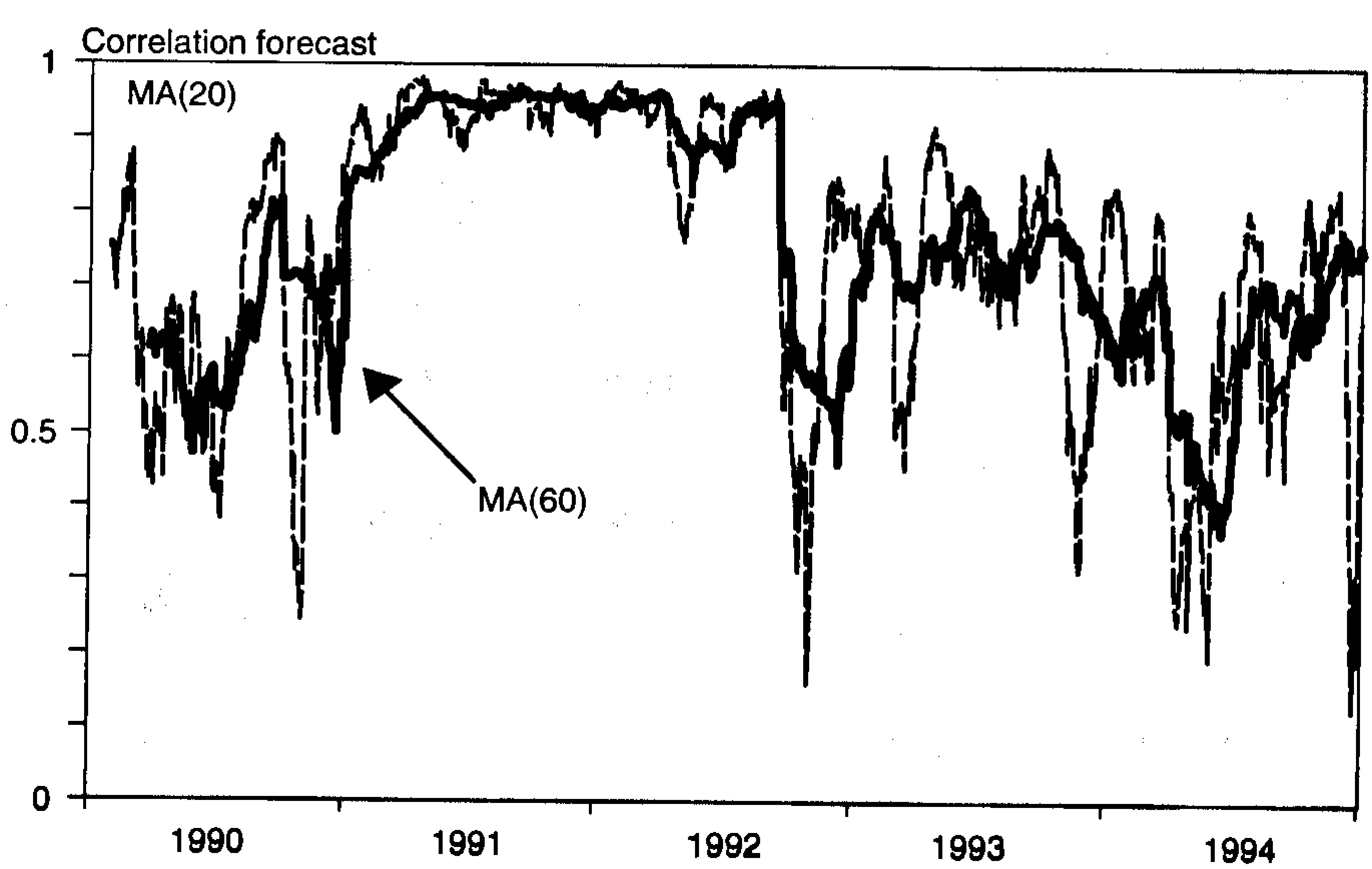
One advantage of multivariate volatility models is that they provide internally consistent risk estimates for a portfolio of assets. Another approach would be to construct the portfolio return series for given weights and to fit a univariate GARCH model to this aggregate series. If the weights change, however, the model has to be estimated again. In contrast, with a multivariate GARCH model, there is no need to reestimate the model for different weights.

9.3.1 Moving Averages

The first method is based on moving averages (MAs), using a fixed window of length M. Figure 9-9 presents estimates based on an MA(20) and MA(60). Correlations start low, at around 0.5, and then increase to 0.9 as the pound enters the EMS. During the September 1992 crisis, correlations

FIGURE 9-9

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



drop sharply and then go back to the pre-EMS pattern. The later drop in correlation would have been disastrous for positions believed to be nearly riskless on the basis of EMS correlations.

These estimates are subject to the same criticisms as before. Moving averages place the same weight on all observations within the moving window and ignore the fact that more recent observations may contain more information than older ones. In addition, dropping observations from the window sometimes has severe effects on the measured correlation.

9.3.2 GARCH

In theory, GARCH estimation could be extended to a multivariate framework. The problem is that the number of parameters to estimate increases exponentially with the number of series.

With two series, for instance, the most general model allows full interactions between each conditional covariance term and the product of lagged innovations and lagged covariances. Expanding Equation (9.2), the first variance term is

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

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

This leads to 7 estimates times 3 series, or 21 parameters. For larger numbers of risk factors, this number quickly becomes unmanageable. This is why simplifications are used often, as shown in Appendix 9.A. Even so, multivariate GARCH systems involve many parameters, which sometimes renders the optimization unstable.

9.3.3 Exponential Averages

Here shines the simplicity of the RiskMetrics approach. Covariances are estimated, much like variances, using an exponential weighing scheme, that is,

$$h_{12,t} = \lambda h_{12,t-1} + (1 - \lambda)\, r_{1,t-1} r_{2,t-1} \tag{9.13}$$

As before, the decay factor λ is arbitrarily set at 0.94 for daily data and 0.97 for monthly data. The conditional correlation then is

$$\rho_{12,t} = \frac{h_{12,t-1}}{\sqrt{h_{1,t-1}h_{2,t-1}}} \tag{9.14}$$

Figure 9-10 displays the time variation in the correlation between the pound and the mark. The pattern of movement in correlations is smoother than in the MA models.

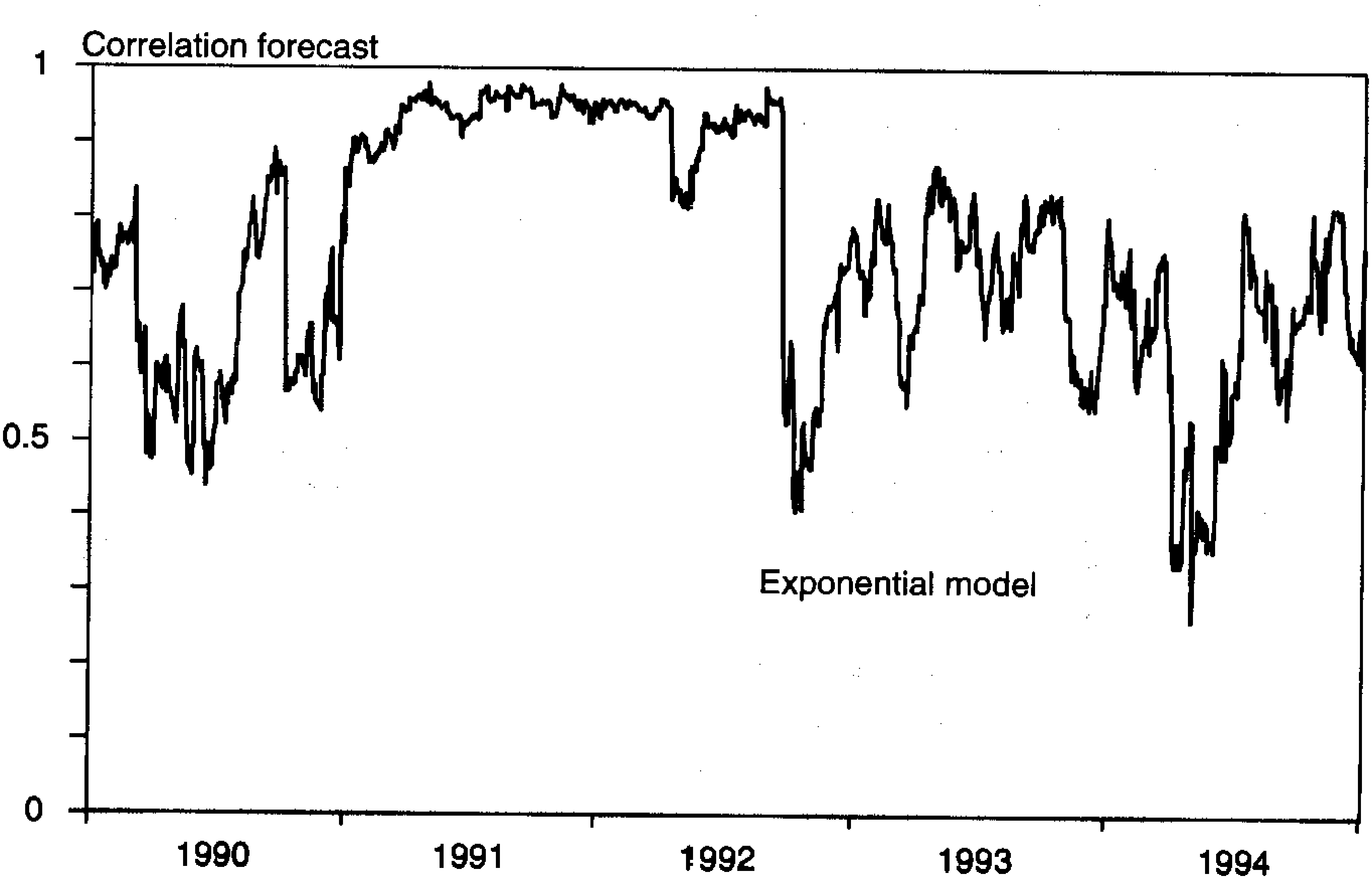
Note that the reason why RiskMetrics sets a common factor λ across all series is to ensure that all estimates of ρ are between -1 and 1. Otherwise, there is no guarantee that this will always be the case.

Even so, this method has a small number of effective observations owing to the rapid decay of weights. The problem is that in order for the covariance matrix to be positive definite, we need at least as many time-series observations as number of assets, as shown in Chapter 8. This explains why the RiskMetrics-provided covariance matrix, with its large number of assets, typically is not positive definite.

By imposing the same decay coefficient for all variances and covariances, this approach is also very restrictive. This reflects the usual tradeoff between parsimony and flexibility.

FIGURE 9-10

Exponential correlation: $/BP and $/DM.



9.3.4 Crashes and Correlations

Low correlations help to reduce portfolio risk. However, it is often argued that correlations increase in periods of global turbulence. If true, such statements are particularly worrisome because increasing correlations occurring at a time of increasing volatility would defeat the diversification properties of portfolios. Measures of VAR based on historical data then would seriously underestimate the actual risk of failure because not only would risk be understated, but so also would correlations. This double blow could well lead to returns that are way outside the range of forecasts.

Indeed, we expect the structure of the correlation matrix to depend on the type of shocks affecting the economy. Global factors, such as the oil crises and the Gulf War, create increased turbulence and increased correlations. Longin and Solnik (1995), for instance, examine the behavior of correlations of national stock markets and find that correlations typically increase by 0.12 (from 0.43 to 0.55) in periods of high turbulence. Recall from Section 7.1 that the risk of a well-diversified portfolio tends to be proportional to $\sqrt{\rho}$. This implies that VAR should be multiplied by a factor proportional to the square root of (0.55/0.43), or 1.13. Thus, just because of the correlation effect, VAR measures could underestimate true risk by 13 percent. Another interpretation of this changing correlation is that the relationship between these risk factors is more complex than the usual multivariate normal distribution and should be modeled with a copula that has greater dependencies in the tail, as seen in Chapter 8.

The extent of bias, however, depends on the sign of positions. Higher correlations are harmful to portfolios with only long positions, as is typical of equity portfolios. In contrast, decreasing correlations are dangerous for portfolios with short sales. Consider our previous example where a trader is long pounds and short marks. As Figure 9-4 shows, this position would have been nearly riskless in 1991 and in the first half of 1992, but the trader would have been caught short by the September 1992 devaluation of the pound. Estimates of VAR based on the previous year's data would have grossly underestimated the risk of the position.

Perhaps these discomforting results explain why regulators impose large multiplicative factors on internally computed VAR measures. But these observations also point to the need for stress simulations to assess the robustness of VAR measures to changes in correlations.

9.4 USING OPTIONS DATA

Measures of VAR are only as good as the quality of forecasts of risk and correlations. Historical data, however, may not provide the best available forecasts of future risks. Situations involving changes in regimes, for instance, are simply not reflected in recent historical data. This is why it is useful to turn to forecasts implied in options data.

9.4.1. Implied Volatilities

An important function of derivatives markets is *price discovery*. Derivatives provide information about market-clearing prices, which includes the discovery of volatility. Options are assets whose price is influenced by a number of factors, all of which are observable save for the volatility of the underlying price. By setting the market price of an option equal to its model value, one can recover an *implied volatility*, or implied standard deviation (ISD).[8] Essentially, the method consists of inverting the option pricing formula, finding σ_{ISD} that equates the model price f to the market price, given current market data and option features, that is,

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

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

This approach can be used to infer a term structure of ISDs every day, plotting the ISD against the maturity of the associated option. Note that σ_{ISD} corresponds to the *average* volatility over the life of the option instead of the instantaneous, overnight volatility. If quotes are available only for longer-term options, we will need to extrapolate the volatility surface to the near term.

Implied correlations also can be recovered from triplets of options on the same three assets. Correlations are also implicit in so-called quanto options, which involve two random variables. An example of a quantity-adjusted option, for instance, would be an option struck on a foreign stock

[8] One potential objection to the use of option volatilities is that the Black-Scholes (BS) model is, *stricto sensu*, inconsistent with stochastic volatilities. Recent research on the effect of stochastic volatilities, however, has shown that the BS model performs well for short-term at-the-money options. For other types of options, such as deep out-of-the-money options, the model may be less appropriate, creating discrepancies in implied volatilities known as the *volatility smile*. For further details, see Bates (1995), Heston (1993), and Duan (1995).

index where the foreign currency payoff is translated into dollars at a fixed rate. The valuation formula for such an option also involves the correlation between two sources of risk. Thus options potentially can reveal a wealth of information about future risks and correlations.

These observations should be tempered with a word of warning. Option ISDs are really for *risk-neutral* (RN) distributions. In fact, we require an estimate of volatility for the *actual*, or physical, distribution. A systematic bias could be introduced between the RN volatility and the actual volatility forecast, reflecting a risk premium. Thus the ISD could be systematically too high relative to the actual volatility, perhaps reflecting investor demand for options, pushing up the ISDs. As long as the difference is constant, however, time variation in the option ISD should provide useful information for time variation in actual risk.

9.4.2 ISDs as Risk Forecasts

If options markets are efficient, the ISD should provide the market's best estimate of future volatility. After all, options trading involves taking volatility bets. Expressing a view on volatility has become so pervasive in the options markets that prices are often quoted in terms of bid-ask volatility. Since options reflect the market consensus about future volatility, there are sound reasons to believe that options-based forecasts should be superior to historical estimates.

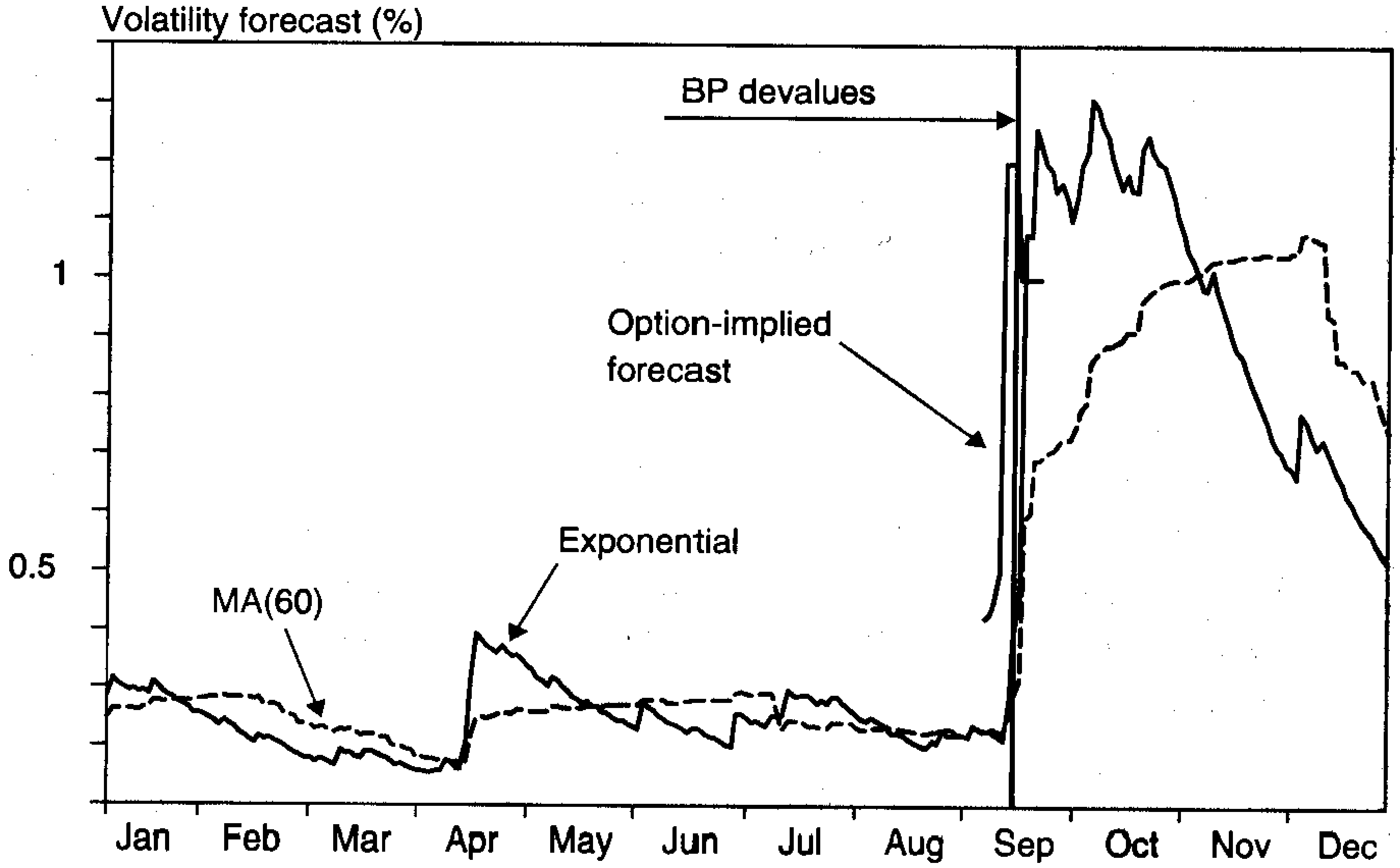
The empirical evidence indeed points to the superiority of options data.[9] An intuitive way to demonstrate the usefulness of options data is to analyze the September 1992 breakdown of the EMS. Figure 9-11 compares volatility forecasts during 1992, including that implied from DM/BP cross-options, the RiskMetrics volatility, and a moving average with a window of 60 days.

As sterling came under heavy selling pressures by speculators, the ISD moved up sharply, anticipating a large jump in the exchange rate. Indeed, sterling went off the EMS on September 16. In contrast, the RiskMetrics volatility only moved up *after* the first big move, and the MA volatility changed ever so slowly. Since options traders rationally anticipated greater turbulence, the implied volatility was much more useful than time-series models.

[9] Jorion (1995a), for instance, shows that for currency futures, options-implied volatilities subsume all information contained in time-series models. Campa and Chang (1998) find that the implied correlation for the dollar/mark and dollar/yen rates outperforms all historical models.

FIGURE 9-11

Volatility forecasts: DM/pound.



Overall, the evidence is that options contain a wealth of information about price risk that is generally superior to time-series models. This information is particularly useful in times of stress, when the market has access to current information that is simply not reflected in historical models. Therefore, my advice is as follows: *Whenever possible, VAR should use implied parameters.*

The only drawback to options-implied parameters is that the menu of traded options is not sufficiently wide to recover the volatility of all essential financial prices. Even fewer cross-options could be used to derive implied correlations. Since more and more options contracts and exchanges are springing up all over the world, however, we will be able to use truly forward-looking options data to measure risk. In the meantime, historical data provide a useful alternative.

9.5 CONCLUSIONS

Modeling time variation in risk is of central importance for the measurement of VAR. This chapter has shown that for most financial assets, short-term volatility varies in a predictable fashion. This variation can be modeled using time-series models such as moving average, GARCH, and exponential weights. These models adapt with varying speeds to changing conditions in financial markets.

The drawback of historical models, unfortunately, is that they are always one step too late, starting to react *after* a big movement has occurred. For some purposes, this is insufficient, which is why volatility forecasts ideally should use information in options values, which are forward-looking.

Finally, it should be noted that GARCH models will induce a lot of movement in 1-day VAR forecasts. While this provides a more accurate forecast of risk over the next day, this approach is less useful for setting risk limits and capital charges.

Assume, for example, that a trader has a VAR risk limit based on a 1-day GARCH model and that the position starts slightly below the VAR limit. A large movement in the market risk factor then will increase the GARCH volatility, thereby increasing the VAR of the actual position that could well exceed the VAR limit. Normally, the position should be cut to decrease the VAR below its limit. The trader, however, will protest that the position has not changed and that this spike in volatility is temporary anyway.

Similarly, the VAR model should not be too volatile if capital charges are based on VAR. Capital charges are supposed to absorb a large shock over a long horizon. Using a 1-day GARCH volatility and the square-root-of-time rule will create too much fluctuation in the capital charge. In such situations, slow-moving volatility models are more appropriate.

Multivariate GARCH models are also ill suited to large-scale risk management problems, which involve a large number of risk factors. This is so because there are simply too many parameters to estimate, unless drastic simplifications are allowed. Perhaps this explains why in practice few institutions use such models at the highest level of aggregation.

APPENDIX 9.A

Multivariate GARCH Models

Multivariate GARCH processes are designed to model time variation in the full covariance matrix. The main issue is that the dimensionality of the model increases very quickly with the number of series N unless simplifications are adopted. Consider, for example, a two-variable system. The covariance matrix has $M = N(N+1)/2 = 3$ entries. This number grows at the speed of N^2 as N increases.

The first class of models generalizes univariate GARCH models. This leads to the VEC(1,1) model, defined as

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

In matrix notation, this is

$$h_t = c + A\eta_{t-1} + Bh_{t-1} \tag{9.17}$$

Note that h_t is a vector with stacked values of variances and covariances, so this is called the *vector* (VEC) *model*. This involves, however, 21 parameters. In general, this number is $N(N+1)/2 + 2\,[N(N+1)/2]^2$, which grows very quickly with N. For $N = 3$, this is already 78. This is too many to be practical.

The first simplification consists of assuming a diagonal matrix for both A and B. This model, called *diagonal VEC* (DVEC), reduces the number of parameters to 9 when $N = 2$. An even simpler version, called the *scalar model*, constrains the matrices A and B to be a positive scalar times a matrix of ones. RiskMetrics is a particular case of the scalar model, where $c = 0$, $a = 1-\lambda$, and $b = \lambda$. The issue is whether imposing the same dynamics on every component is a reasonable assumption.

Generally, a major problem with multivariate GARCH models is that the resulting covariance matrix H_t must be positive definite at every point in time. This could be achieved by imposing restrictions on the parameters, but in practice this is difficult to enforce.

One way to ensure positive definiteness is to use a parametrization proposed by Baba, Engle, Kraft, and Kroner (BEKK) (1990). The BEKK model is

$$H_t = C'C + A'r_{t-1}r'_{t-1}A + B'H_{t-1}B \tag{9.18}$$

where C, A, and B are $(N \times N)$ matrices, but C is upper triangular, with zeroes below the diagonal. This is a special case of the VEC model. The number of parameters is $3 + 4 + 4 = 11$, which is indeed fewer than that of the VEC model. In general, this number is $N(N + 1)/2 + N^2 + N^2$. To simplify further, one could impose diagonal matrices A and B, which is a special case of the DVEC model, or force the matrices to be proportional to a scalar a and b.

A particular case of the BEKK model is the *factor model*, which assumes that the time variation is driven by a small number of factors, $g_{1,t}, \ldots, g_{K,t}$, each following a GARCH(1,1) process. The one-factor model is

$$H_t = C'C + b_1 b_1' g_{1,t} \tag{9.19}$$

where the variance factor is modeled as

$$g_{1,t} = 1 + \alpha_1 f_{1,t}^2 + \beta_1 g_{1,t-1} \tag{9.20}$$

and the factor f_t can be specified as a linear function of r_t. The number of parameters is now reduced to $3 + 2 + 2 = 7$. In general, this number is $N(N + 1)/2 + N + 2$.

Another class of models consists of nonlinear combinations of univariate GARCH models. Each series is modeled individually first. The variance forecasts then are combined with a correlation structure. For instance, the *constant conditional correlation* (CCC) *model* imposes fixed correlations. This is

$$H_t = D_t R D_t = \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\ 0 & \sqrt{h_{22,t}} \end{bmatrix} \tag{9.21}$$

where each entry has the form $\rho_{ij}\sqrt{h_{ii,t}h_{jj,t}}$. This contains $1 + 3 + 3 = 7$ parameters. In general, this number is $N(N - 1)/2 + 3N$. Of course, the assumption of constant conditional correlations may appear unrealistic. The alternative is a *dynamic conditional correlation model* (DCC). Engle (2002) expands Equation (9.21) to a time-varying correlation matrix R_t, that is,

$$R_t = \begin{bmatrix} 1/\sqrt{q_{11,t}} & 0 \\ 0 & 1/\sqrt{q_{22,t}} \end{bmatrix} \begin{bmatrix} q_{11,t} & q_{12,t} \\ q_{12,t} & q_{22,t} \end{bmatrix} \begin{bmatrix} 1/\sqrt{q_{11,t}} & 0 \\ 0 & 1/\sqrt{q_{22,t}} \end{bmatrix} \tag{9.22}$$

where the $(N \times N)$ symmetric matrix Q_t follows a GARCH-type process, that is,

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha\epsilon_{t-1}\epsilon'_{t-1} + \beta Q_{t-1} \tag{9.23}$$

with ϵ_t defined as the vector of scaled residuals. $\bar{Q}$ is set to the unconditional covariance matrix. Because α and β are scalars, all conditional correlations obey the same dynamics. This, however, ensures that the correlation matrix R_t is positive definite. This model contains $7 + 2 = 9$ parameters when $N = 2$. In general, this number is $N(N - 1)/2 + 3N + 2$ when there is one common factor only.

Overall, the main issue in multivariate GARCH modeling is to provide a realistic but still parsimonious representation of the covariance matrix. The models presented here cut down the number of parameters considerably. For a detailed review of this very recent and quickly expanding literature, interested readers should see Bauwens et al. (2005).

QUESTIONS

1. In practice, we seem to observe too many extreme observations than warranted by the normal distribution. Give two explanations for this observation.
2. The moving average is one approach to estimate volatility. List two drawbacks to this method.
3. Which volatility forecast is more volatile and why? An MA process with a window of 20 days or 60 days?
4. In the GARCH(1,1) process $h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1}$, what is the unconditional variance?
5. What is the restriction on the sum of the parameters for the GARCH (1,1) model to be stationary?

6. Why can the exponential weighted-moving-average (EWMA) approach be viewed as a special case of the GARCH process?
7. The GARCH model assumes that the scaled residual $\epsilon_t = r_t/\sqrt{h_t}$ follows a conditional normal distribution. How can this model be extended to both time variation in volatility and conditional fat tails?
8. Assume that a risk manager uses a simple square root of time to extrapolate the variance to 10 days. In reality, the process is a GARCH model and starts with current variance above the long-term average. Will the simple rule overestimate or underestimate risk?
9. Assume that the decay factor is chosen as $\lambda = 0.94$ for the EWMA model with daily data. What is the weight on the latest observation and on that of the day before?
10. For the EWMA model with decay of 0.94, the number of effective observations is said to be small. Explain.
11. The current estimate of daily volatility is 1 percent. The latest return is 2 percent. Using the EWMA model with $\lambda = 0.94$, compute the updated estimate of volatility.
12. Continue with the preceding question. As of now, what is the volatility forecast for the following day, $t + 1$?
13. The RiskMetrics approach uses the EWMA model with decay of 0.94 for daily data and 0.97 for monthly data. Why is this inconsistent?
14. Why is the general GARCH model not used commonly to model the full covariance matrix?
15. Why is the EWMA with the same decay convenient for modeling the full covariance matrix?
16. Explain why we need to bother about modeling the *joint* distribution of N risk factors. Given the problems created by the dimensionality that increases with the square of N, it would seem simpler to apply *univariate GARCH* to the current portfolio only.
17. Under what situations are historical models not a good measure of volatility?
18. What is the advantage of using ISD (implied standard deviation) to predict volatility?

PART III

VALUE-AT-RISK SYSTEMS

CHAPTER 10

VAR Methods

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

—*Attributed to a French mathematician*

Value at risk (VAR) has become an essential tool for risk managers because it provides a quantitative measure of downside risk based on current positions. In practice, the objective should be to provide a reasonably accurate estimate of risk at a reasonable cost. This involves choosing from among the various industry standards a method that is most appropriate for the portfolio at hand. To help with this selection, this chapter presents and critically evaluates various approaches to VAR.

The potential for losses results from exposures to the risk factors, as well as the distributions of these risk factors. This dichotomy finds its way into the structure of risk management systems, which can be classified into models for exposure and models for the distributions of risk factors.

Models for exposure can be classified into two groups. The first group uses local valuation. *Local-valuation methods* measure risk by valuing the portfolio once, at the initial position, and using local derivatives to infer possible movements. Within this class, the *delta-normal method* uses linear, or delta, exposures and assumes normal distributions. This is sometimes called the *variance-covariance method.* For portfolios exposed to a small number of risk factors, second-order derivatives sometimes are used. The second group uses full valuation. *Full-valuation methods* measure risk by fully repricing the portfolio over a range of scenarios.

Models for risk factors include parametric approaches, such as the normal distribution, and nonparametric approaches based on historical data.

Section 10.1 gives an overview of VAR systems. The local- and full-valuation approaches are discussed in Section 10.2. Initially, we consider a simple portfolio that is driven by one risk factor only. This chapter then turns to VAR methods for large portfolios. The delta-normal method is explained in Section 10.3. The historical simulation and Monte Carlo (MC) simulation methods are discussed next in Sections 10.4 and 10.5. All these methods require mapping, which is developed in Chapter 11.

This classification reflects a fundamental tradeoff between speed and accuracy. Speed is important for large portfolios exposed to many risk factors that involve a large number of correlations. These are handled most easily in the delta-normal approach. Accuracy may be more important, however, when the portfolio has substantial nonlinear components. Section 10.6 presents some empirical comparisons of the VAR approaches. Finally, Section 10.7 summarizes the pros and cons of each of the three main methods.

10.1 VAR SYSTEMS

The potential for gains and losses can be attributed to two sources. On the one hand are the exposures, which represent active choices by the trader or portfolio manager. On the other hand are the movements in the risk factors, which are outside their control.

This dichotomy is reflected in the structure of risk management systems, which is described in Figure 10-1. The left-hand side describes the portfolio *positions*, which have as input trades from the front office. The right-hand side describes the *risk factors*, which have as input data feeds with current market prices. Positions and risk-factor distributions are brought together in the *risk engine*, which generates a distribution of portfolio values that can be summarized, for instance, by its VAR.

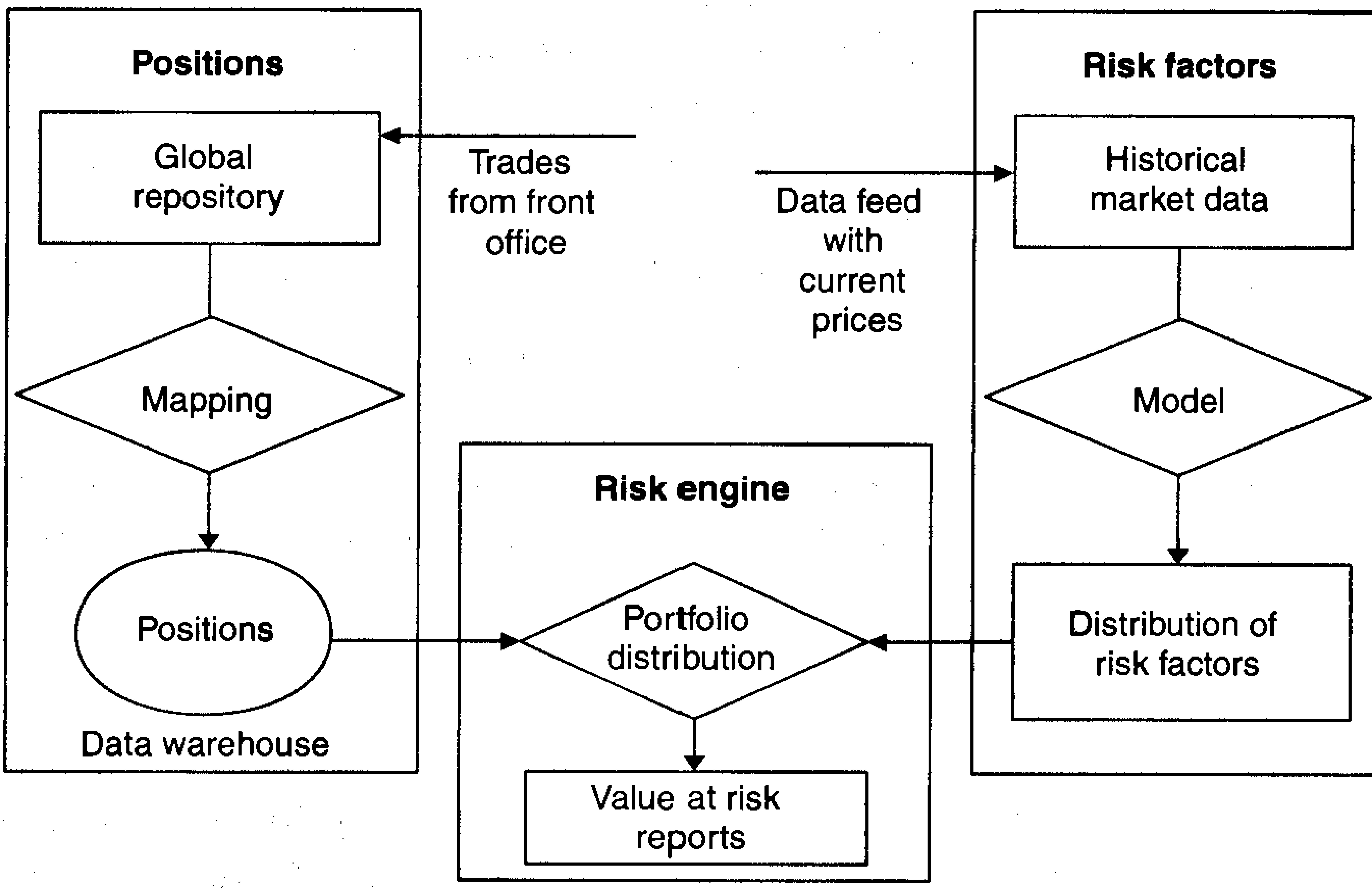
Different VAR methods make different assumptions for the modeling of positions and risk factors. The positions can be replaced by their linear exposures to the risk factors, or by the quadratic exposures, or by using full repricing. The distributions of risk factors can be modeled using a normal distribution, or the historical data, or Monte Carlo simulations.

Modern risk measurement methods are applied at the highest level of the portfolio. This generally involves a very large number of instruments and risk factors. It would be impractical to model all these positions individually. Realistically, simplifications are required.

The first step in the implementation of a risk measurement system involves choosing an appropriate number of risk factors. Positions then are simplified by *mapping* each and every one on the risk factors. This

FIGURE 10-1

VAR systems.



replaces the dollar value of positions in each instrument by a set of dollar exposures on the risk factors. These exposures then are aggregated across the whole portfolio to create net positions that are matched to the risk factors. This mapping process will be detailed further in Chapter 11. In this chapter we focus on integration of exposures with the risk factors.

10.2 LOCAL VERSUS FULL VALUATION

10.2.1 Delta-Normal Valuation

Local-valuation methods measure exposures with partial derivatives. To illustrate the approach, take an instrument whose value depends on a single underlying risk factor S. The first step consists of valuing the asset at the initial point, that is,

$$V_0 = V(S_0) \tag{10.1}$$

along with analytical or numerical derivatives. Define delta (Δ_0) as the first partial derivative, or the asset sensitivity to changes in prices, evaluated at the current position V_0. This would be called *delta* for a derivative or *modified duration* for a fixed-income portfolio. For instance, with an at-the-money call, $\Delta = 0.5$, and a long position in one option is simply replaced by a 50 percent position in one unit of the underlying asset. Thus this is a linear exposure to the risk factor.

The potential loss in value of an option dV then is computed as

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

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

Because this is a linear relationship, the worst loss for V is attained for an extreme value of S. If the distribution is normal, the portfolio VAR can be derived from the product of the exposure and the VAR of the underlying variable, that is,

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

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

This approach is called the *delta-normal method.* Because VAR is obtained as a closed-form solution, this is an *analytical* method. Note that VAR was measured by computing the portfolio value only once, at the current position V_0.[1]

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

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

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

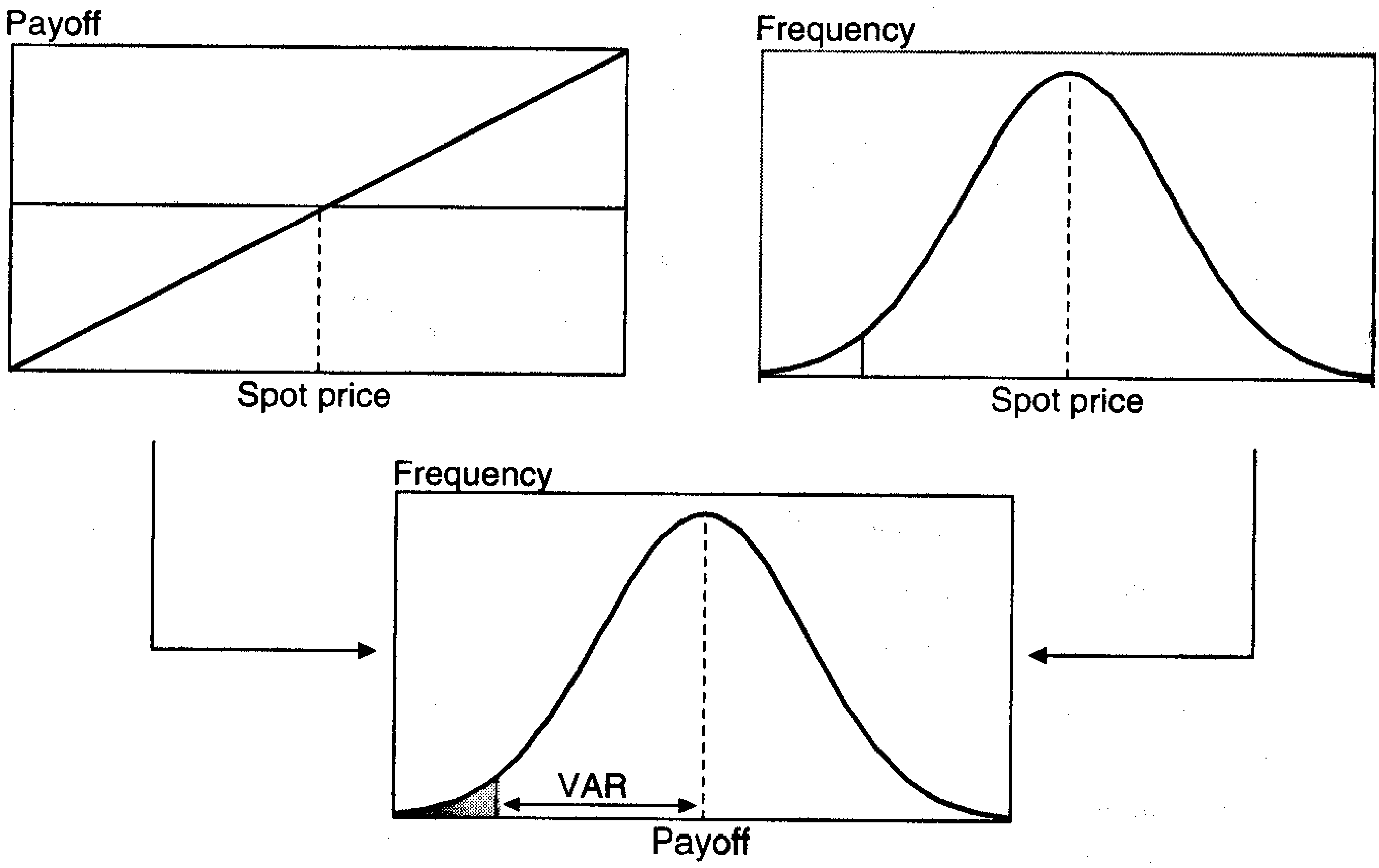
$$\text{VAR} = |D^*V| \times (\alpha\sigma) \tag{10.5}$$

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

[1] Usually, delta can be computed easily. If the instrument is an option valued using a binomial tree, for instance, delta can be computed from the up and down values of the option at the first step divided by the changes in asset values.

FIGURE 10-2

Distribution with linear exposures.



This method is illustrated in Figure 10-2, where the profit payoff V is a linear function of the underlying spot price S and is displayed in the upper left panel. As shown in the right panel, the price is normally distributed. As a result, the profit itself is normally distributed, as shown at the bottom of the figure. This implies that the VAR for the profit can be derived from the exposure and the VAR for the underlying price. There is a one-to-one mapping between the two VAR measures.

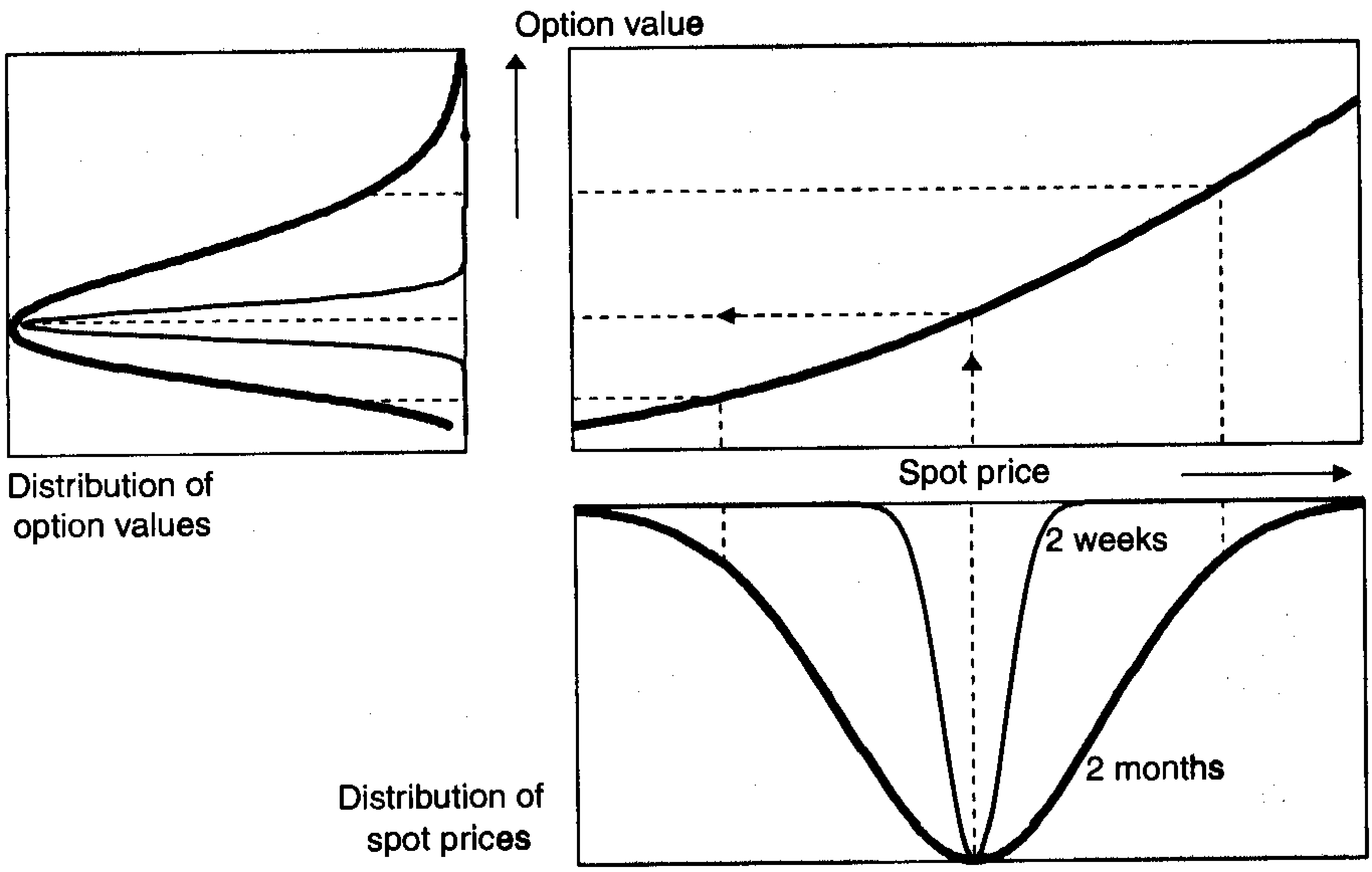
10.2.2 Full Valuation

In some situations, the delta-normal approach is totally inadequate. This is the case, for instance, with combinations of options that have very nonlinear payoffs. Sometimes the worst loss may not be obtained for extreme realizations of the underlying spot rate.

Consider, for instance, a simple example of a long position in a call option. In this case, we can describe the distribution of option values easily. This is so because there is a one-to-one relationship between V and S. In other words, given the pricing function, any value for S can be translated into a value for V, and vice versa.

FIGURE 10-3

Transformation of distributions.



This is illustrated in Figure 10-3, which shows how the distribution of the spot price is translated into a distribution for the option value (in the left panel). Note that the option distribution has a long right tail owing to the upside potential, whereas the downside is limited to the option premium. This shift is due to the nonlinear payoff on the option. Note how the positive skewness translates into a shorter left tail or lower VAR than otherwise.

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

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

Because this is a *monotonic* transformation, the quantiles can be translated from S to V directly. This result, unfortunately, does not translate to general payoff functions.

The nonlinearity effect is not obvious, though. It also depends on the maturity of the option and on the range of spot prices over the horizon. The option illustrated here is a call option with 3 months to expiration. To obtain a visible shift in the shape of the option distribution, the volatility was set at 20 percent per annum and the VAR horizon at 2 months, which is rather long.

The figure also shows thinner distributions that correspond to a VAR horizon of 2 weeks. Here, the option distribution is indistinguishable from the normal. In other words, the mere presence of options does not necessarily invalidate the delta-normal approach. The quality of the approximation depends on the extent of nonlinearities, which is a function of the type of options, of their maturities, as well as of the volatility of risk factors and VAR horizon. The shorter the VAR horizon, the better is the delta-normal approximation.

Equation (10.6) is a convenient transformation of quantiles but does not apply with more complex, nonmonotonic payoffs. An example is that of a long *straddle*, which involves the purchase of a call and a put. The worst payoff, which is the sum of the premiums, will be realized if the spot rate does not move at all. In general, it is not sufficient to evaluate the portfolio at the two extremes. All intermediate values must be checked.

The *full-valuation approach* considers the portfolio value for a wide range of price levels, that is,

$$dV = V(S_1) - V(S_0) \tag{10.7}$$

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

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

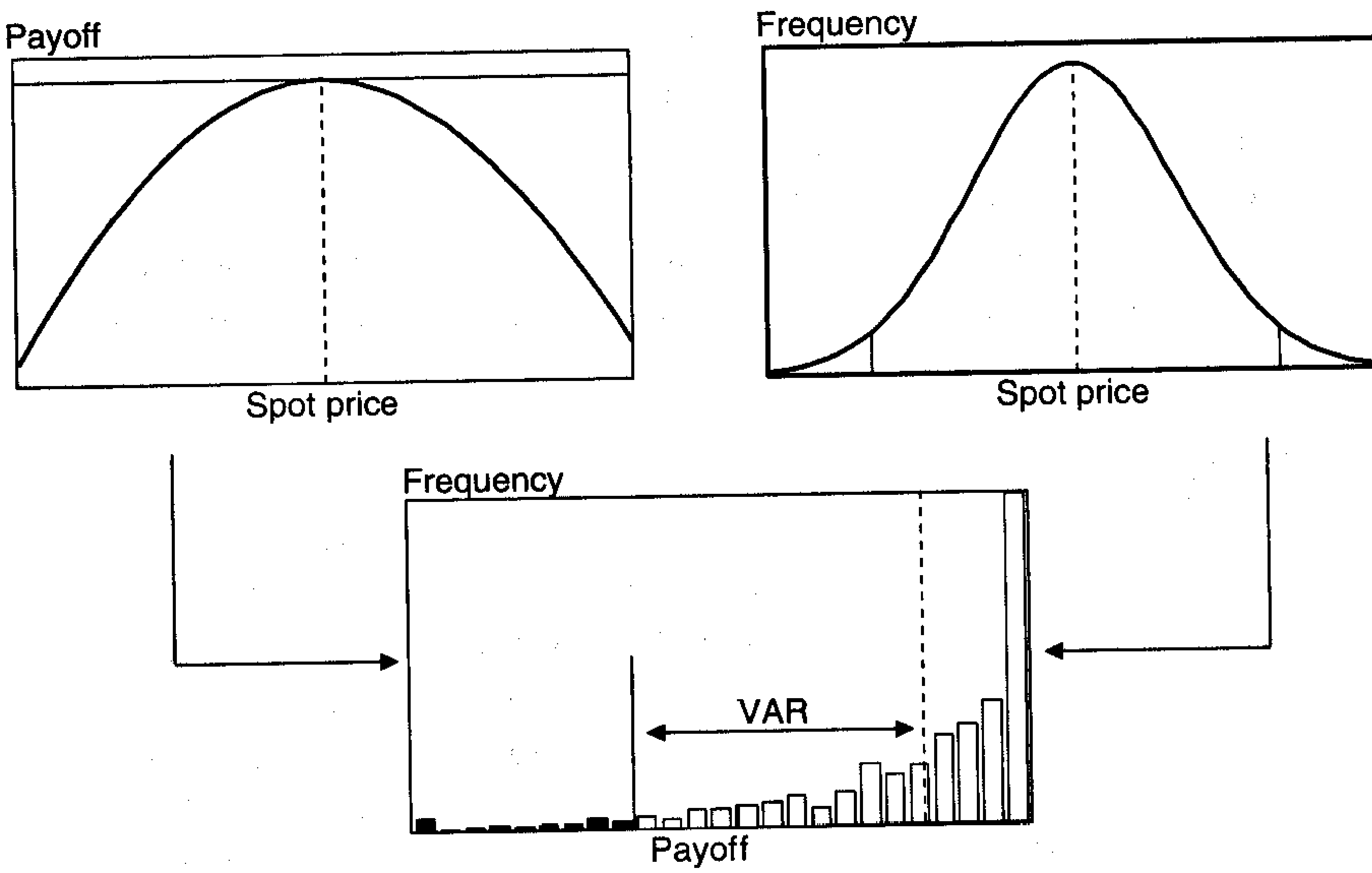
Alternatively, the *historical simulation approach* simply samples from recent historical data.

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

To illustrate the result of nonlinear exposures, Figure 10-4 displays the payoff function for a short straddle that is highly nonlinear. The resulting distribution is severely skewed to the left. Further, there is no direct way to relate the VAR of the portfolio to that of the underlying asset.

FIGURE 10-4

Distribution with nonlinear exposures.



Computationally, this approach is quite demanding because it requires marking to market the whole portfolio over a large number of realizations of underlying random variables. As a result, methods have been developed to speed up the computations. In general, these approaches try to break the link between the number of Monte Carlo draws and the number of times the portfolio is repriced.

One example is the *grid Monte Carlo approach*, which starts by an exact valuation of the portfolio over a limited number of grid points.[2] For each simulation, the portfolio value then is approximated using a linear interpolation from the exact values at the adjoining grid points. This approach is especially efficient if exact valuation of the instrument is complex. Take, for instance, a portfolio with one risk factor for which we require 1000 values $V(S_1)$. With the grid Monte Carlo method, 10 full valuations at the grid points may be sufficient. In contrast, the full Monte Carlo method would require 1000 full valuations.

[2] Picoult (1997) describes this method in more detail.

10.2.3 Delta-Gamma Approximations (The "Greeks")

It may be possible to extend the analytical tractability of the delta-normal method with higher-order terms. Because the method uses partial derivatives defined using Greek letters, it is sometimes called the *Greeks*.

We can improve the quality of the linear approximation by adding terms in the Taylor expansion of the valuation function, that is,

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

where Γ is now the second derivative of the portfolio value, and Θ is the time drift, which is deterministic. For a fixed-income portfolio, the instantaneous price-yield relationship is now

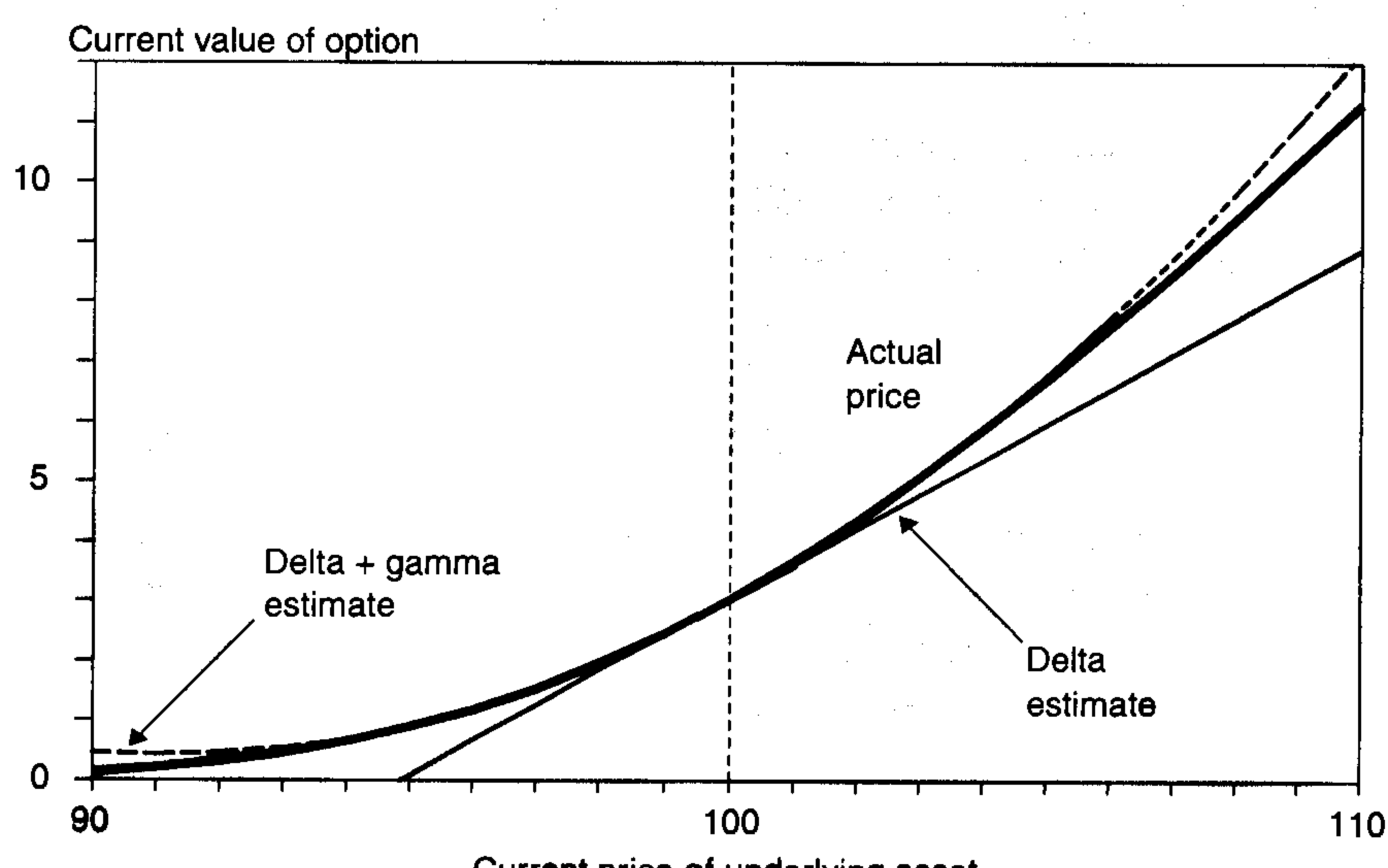
$$dV = -(D^*V)dy + \tfrac{1}{2}(CV)dy^2 + \cdots \qquad (10.10)$$

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

Figure 10-5 describes the approximation for a simple position, a long position in a European call option. The actual price is represented by the

FIGURE 10-5

Delta-gamma approximation for a long call.



thick line. The delta estimate is the straight line below. The delta plus gamma estimate is the dashed line. Because Γ is positive, the term ΓdS^2 must be positive, and the quadratic estimate must lie above the linear estimate. The graph shows that the linear estimate is only valid for small movements around the initial value. For larger movements, the delta-gamma estimate creates a better fit.

The figure also shows that delta is not constant but rather changes as a function of the spot price; gamma gives the rate of change in delta. Delta also changes with the passage of time. This has implications for the extrapolation of risk across horizons. With linear models, as we have seen in Chapter 4, daily VAR can be adjusted easily to other periods by scaling by a square-root-of-time factor. This adjustment assumes that the position is fixed and that daily returns are independent and identically distributed. Such adjustment, however, is not appropriate for options, even when the positions are fixed. This is so because the option delta changes dynamically over time. Hence the square-root-of-time adjustment may not be valid for options.

We now turn to the computation of VAR for the long-call option position. Using the Taylor expansion in Equation (10.6) gives

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

This formula is valid for long and short positions in calls and puts and, more generally, for portfolios whose payoff is monotonic in S. If Γ is positive, which corresponds to a net long position in options, the second term will decrease the linear VAR. Indeed, Figure 10-5 shows that the downside risk for the option is less than that given by the delta approximation. If Γ is negative, which corresponds to a net short position in options, VAR is increased.

This closed-form solution does not apply, unfortunately, to more complex functions $V(S)$. Appendix 10.A lists some analytical approximations, including the *delta-gamma-delta*. Generally, however, quadratic approximations are not used at the highest level of aggregation for large portfolios. Full implementation would require knowledge of all gammas and cross-gammas, that is, second derivatives with respect to other risk factors.

On the other hand, quadratic approximations are very useful to speed up computations with simulations. An example is the *delta-gamma-Monte-Carlo approach*, which creates random simulations of the risk factor S and then uses the Taylor approximation to create simulated movements in the option value. This method is also known as a *partial-simulation approach*. Note that this is still a local-valuation method because the asset is fully valued at the initial point V_0 only. The portfolio can be valued by adding the approximated option positions to all others.

10.2.4 Comparison of Methods

To summarize, Table 10-1 classifies the various VAR methods. Overall, each of these methods is best adapted to a different environment:

- For large portfolios where optionality is not a dominant factor, the delta-normal method provides a fast and efficient method for measuring VAR.
- For fast approximations of option values, mixed methods such as delta-gamma-Monte-Carlo or grid Monte Carlo are efficient.
- For portfolios with substantial option components (such as mortgages) or longer horizons, a full-valuation method may be required.

10.2.5 An Example: Leeson's Straddle

The Barings story provides a good illustration of these various methods. In addition to the long futures positions described in Chapter 7, Leeson also sold options, about 35,000 calls and puts each on Nikkei futures. This

TABLE 10-1

Comparison of VAR Methods

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