CHAPTER 9

Forecasting Risk and Correlations

To have a future in risk management, one needs to include the future in risk measurement.

—*Peter Davies, Askari (a risk management company)*

Chapter 4 described the risk of basic financial variables such as interest rates, exchange rates, and equity prices. A reader looking more closely at the graphs would notice that risk appears to change over time. This is quite obvious for exchange rates, which displayed much more variation after 1973. Bond yields also were more volatile in the early 1980s. These periods corresponded to structural breaks: Exchange rates started to float in 1973, and the Fed abruptly changed monetary policies in October 1979. Even during other periods, volatility seems to *cluster* in a predictable fashion.

The observation that financial market volatility is predictable has important implications for risk management. If volatility increases, so will value at risk (VAR). Investors may want to adjust their portfolio to reduce their exposure to those assets whose volatility is predicted to increase. Also, predictable volatility means that assets depending directly on volatility, such as options, will change in value in a predictable fashion. Finally, in a rational market, equilibrium asset prices will be affected by changes in volatility. Investors who can reliably predict changes in volatility should be able to control financial market risks better.

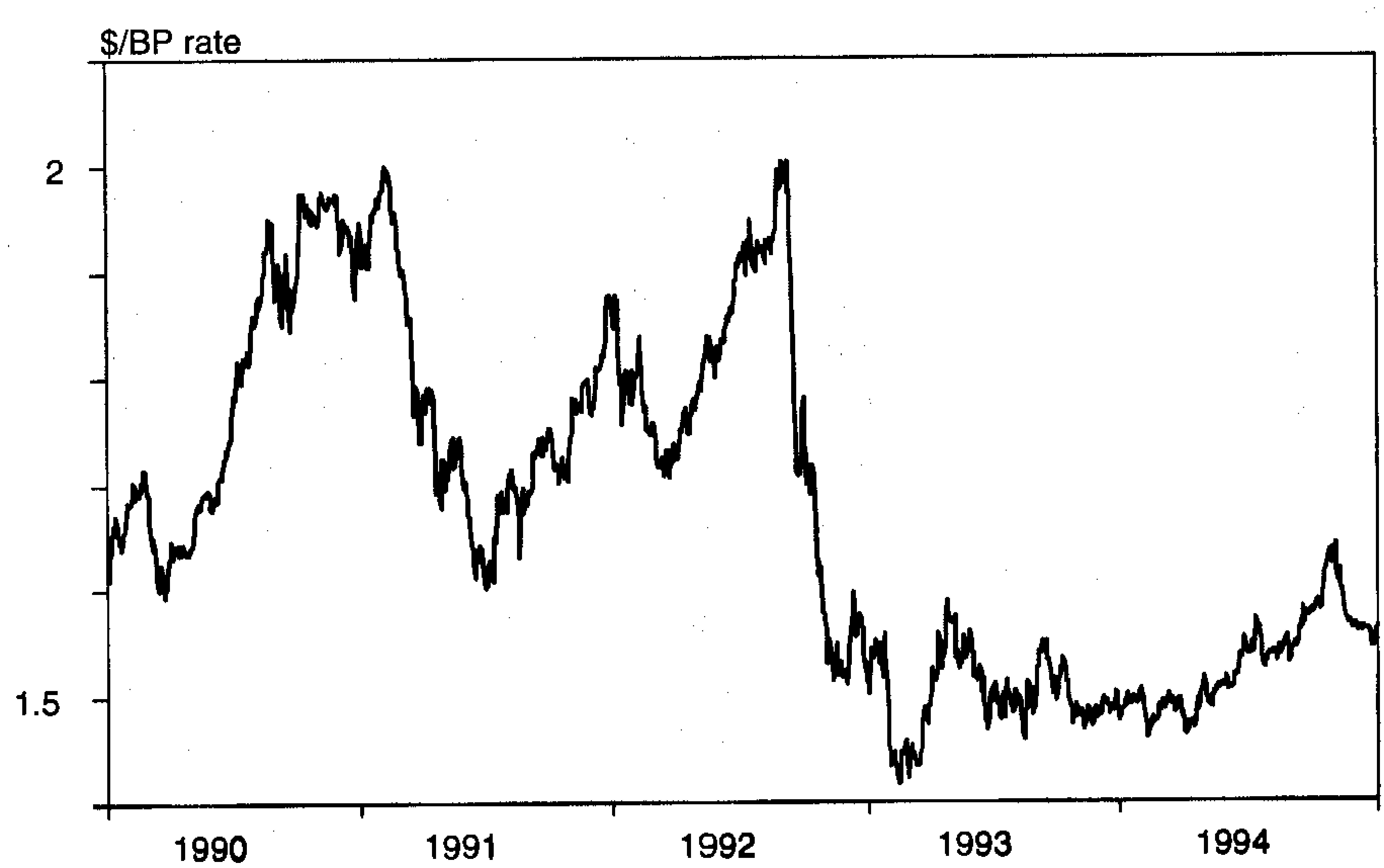
The purpose of this chapter is to present techniques to forecast variation in risk and correlations. Section 9.1 motivates the problem by taking the example of a series that underwent structural changes leading to predictable patterns in volatility. Section 9.2 then presents recent developments in time-series models that capture time variation in volatility. A particular application of these models is the exponential approach adopted for the RiskMetrics system. Section 9.3 extends univariate models to correlation forecasts. Finally, Section 9.4 argues that time-series models are inherently inferior to forecasts of risk contained in options prices.

9.1 TIME-VARYING RISK OR OUTLIERS?

As an illustration, we will walk through this chapter focusing on the U.S. dollar/British pound ($/BP) exchange rate measured at daily intervals. Movements in the exchange rate are displayed in Figure 9-1. The 1990–1994 period was fairly typical, covering narrow trading ranges and wide swings. September 1992 was particularly tumultuous. After vain attempts by the Bank of England to support the pound against the German mark, the pound exited the European Monetary System. There were several days with very large moves. On September 17 alone, the pound fell by 6 percent against the mark and also against the dollar. Hence we can expect interesting patterns in volatility. In particular, the question is whether this structural change led to predictable time variation in risk.

FIGURE 9-1

Spot rate: British pound versus dollar.



Over this period, the average daily volatility was 0.694 percent, which translates into 11.02 percent per annum (using a 252-trading-day adjustment). This risk measure, however, surely was not constant over time. In addition, time variation in risk could explain the fact that the empirical distribution of returns does not quite exactly fit a normal distribution.

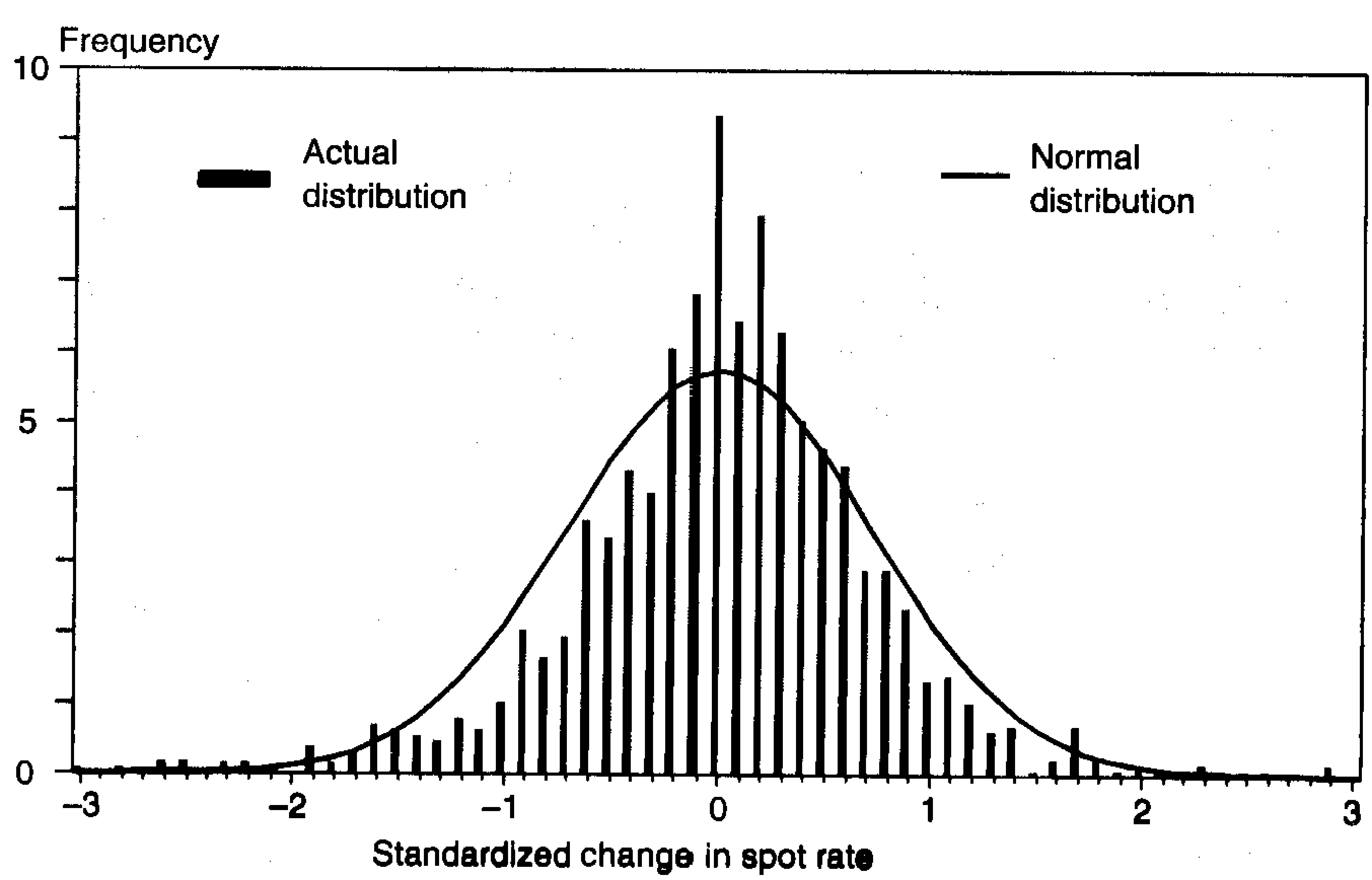
Figure 9-2 compares the normal approximation with the actual empirical distribution of the $/BP exchange rate. Relative to the normal model, the actual distribution contains more observations in the center and in the tails.

These fat tails can be explained by two alternative viewpoints. The first view is that the true distribution is stationary and indeed contains fat tails, in which case a normal approximation is clearly inappropriate. The other view is that the distribution does change through time. As a result, in times of turbulence, a stationary model could view large observations as outliers when they are really drawn from a distribution with temporarily greater dispersion.

In practice, both explanations carry some truth. This is why forecasting variation in risk is particularly fruitful for risk management. In

FIGURE 9-2

Distribution of the $/BP rate.



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

9.2 MODELING TIME-VARYING RISK

9.2.1 Moving Averages

A very crude method, but one that is employed widely, is to use a *moving window* of fixed length for estimating volatility. For instance, a typical length is 20 trading days (about a calendar month) or 60 trading days (about a calendar quarter).

Assuming that we observe returns r_t over M days, this volatility estimate is constructed from a *moving average* (MA), that is,

$$\sigma_t^2 = (1/M)\sum_{i=1}^{M} r_{t-i}^2 \tag{9.1}$$

Here we focus on raw returns instead of returns around the mean. This is so because for most financial series, ignoring expected returns over very short time intervals makes little difference for volatility estimates.

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

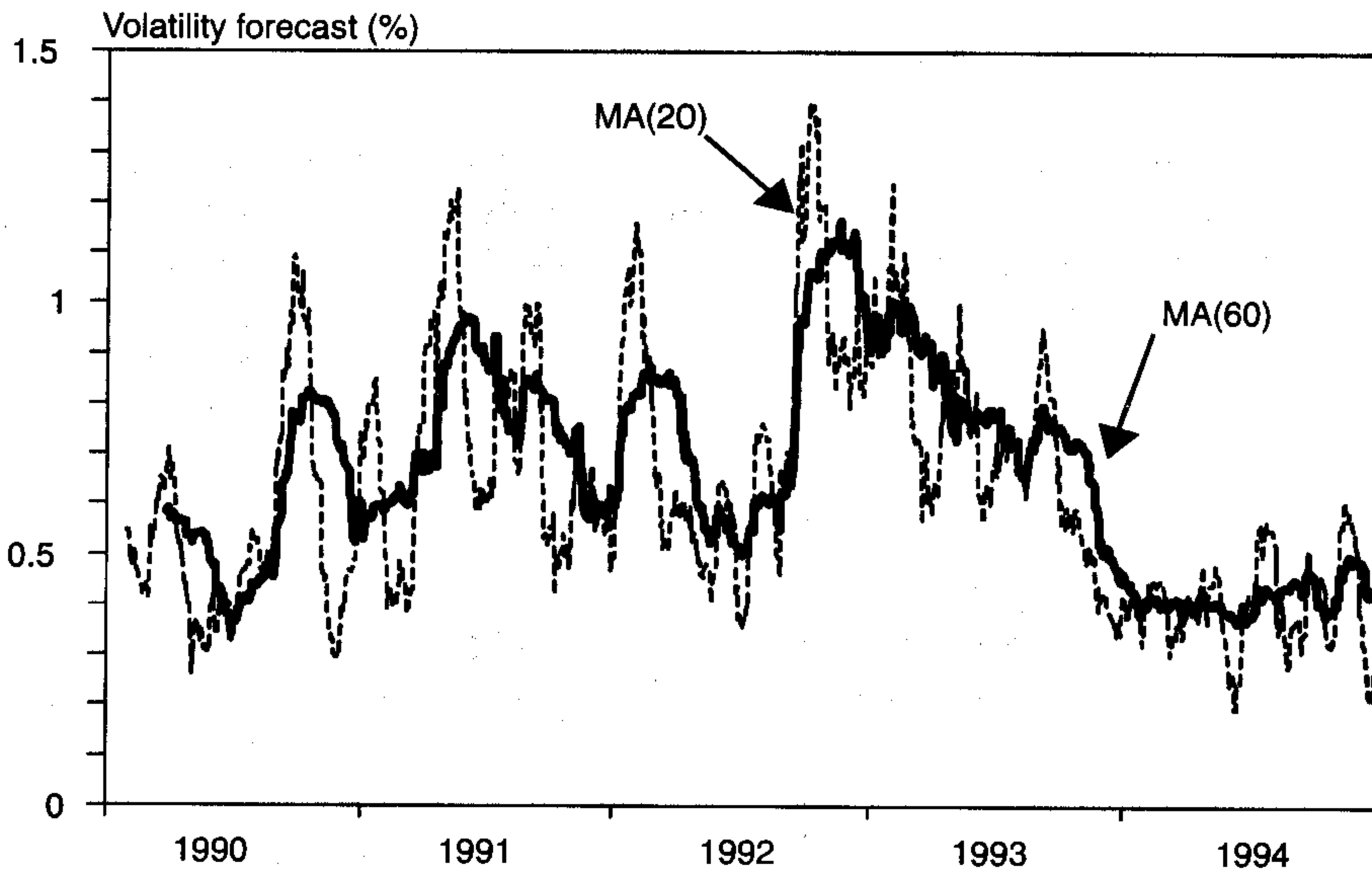
While simple to implement, this model has serious drawbacks. First, it ignores the dynamic ordering of observations. Recent information receives the same weight as older observations in the window that may no longer be relevant.

Also, if there was a large return M days ago, dropping this return as the window moves 1 day forward will affect the volatility estimate substantially. For instance, there was a 3 percent drop on September 17, 1992. This observation will increase the MA forecast immediately, which correctly reflects the higher volatility. The MA(20), however, reverts to a lower value after 20 days; the MA(60) reverts to a lower value after 60 days. As a result, moving-average measures of volatility tend to look like *plateaus* of width M when plotted against time. The subsequent drop, however, is totally an artifact of the window length. This has been called the *ghosting feature* because the MA measure changes for no apparent reason.

[1] Other methods exist, however. Also, risk estimators do not necessarily have to rely solely on daily closing prices. Parkinson (1980) has shown that using the information in the extreme values (daily high and low) leads to an estimator that is twice as efficient as the usual volatility; this is so because it uses more information.

FIGURE 9-3

Moving-average (MA) volatility forecasts.



The figure shows that the MA(60) is much more stable than the MA(20). This is understandable because longer periods decrease the weight of any single day. But is it better? This approach leaves wholly unanswered the choice of the moving window. Longer periods increase the precision of the estimate but could miss underlying variation in volatility.

9.2.2 GARCH Estimation

This is why volatility estimation has moved toward models that put more weight on recent information. The first such model was the *generalized autoregressive conditional heteroskedastic* (GARCH) *model* proposed by Engle (1982) and Bollerslev (1986) (see Box 9-1). *Heteroskedastic* refers to the fact that variances are changing.

The GARCH model assumes that the variance of returns follows a predictable process. The *conditional* variance depends on the latest innovation but also on the previous conditional variance. Define h_t as the conditional variance, using information up to time $t-1$, and r_{t-1} as the previous day's return. The simplest such model is the GARCH(1,1) process, that is,

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1} \qquad (9.2)$$

BOX 9-1

NOBEL RECOGNITION

The importance of measuring time variation in risk was recognized when Professor Robert Engle was awarded the 2003 Nobel Prize in Economics. The Royal Swedish Academy of Sciences stated that Professor Engle's "ARCH models have become indispensable tools not only for researchers but also for analysts on financial markets, who use them in asset pricing and in evaluating portfolio risk."

This announcement was a milestone for the risk management profession because it recognized the pervasive influence of market risk modeling methods.

The average, unconditional variance is found by setting $E(r_{t-1}^2) = h_t = h_{t-1} = h$. Solving for h, we find

$$h = \frac{\alpha_0}{1-\alpha_1-\beta} \tag{9.3}$$

For this model to be stationary, the sum of parameters $\alpha_1+\beta$ must be less than unity. This sum is also called the *persistence*, for reasons that will become clear later on.

The beauty of this specification is that it provides a parsimonious model with few parameters that seems to fit the data quite well.[2] GARCH models have become a mainstay of time-series analysis of financial markets that systematically display volatility clustering. There are literally thousands of papers applying GARCH models to financial series.[3] Econometricians also have frantically created many variants of the GARCH model, most of which provide only marginal improvement on the original model. Readers interested in a comprehensive review of the literature should consult Bollerslev et al. (1992).

The drawback of GARCH models is their nonlinearity. The parameters must be estimated by maximization of the likelihood function, which involves a numerical optimization. Typically, researchers assume that the scaled residuals $\epsilon_t = r_t/\sqrt{h_t}$ have a normal distribution and are independent. If we have T observations, their joint density is the product of the densities

[2] For the theoretical rationale behind the success of GARCH models, see Nelson (1990).

[3] See French et al. (1987) for stock-return data, Engle et al. (1987) for interest-rate data, Hsieh (1988) and Giovannini and Jorion (1989) for foreign-exchange data.

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

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

where f is the normal density function.

In fact, this result is even more general. Bollerslev and Wooldridge (1992) have shown that when the true distribution is not normal, the parameters so estimated are *consistent*.[4] The method is then called *quasi-maximum likelihood*. Thus one could estimate the conditional distribution in two steps, first estimating the GARCH parameters using Equation (9.4) and then estimating the distribution parameters for the *scaled residual*, that is,

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

The conditional distribution of this scaled residual could be taken as a student t or some other parametric distribution or even be sampled from the historical data. The latter approach is called *filtered historical simulation*.

Table 9-1 presents the results of the estimation for a number of financial series over the 1990–1999 period. There are wide differences in the level of volatility across series, yet for all these series, the time variation in risk is highly significant. The persistence parameter is also rather high, on the order of 0.97–0.99, although this depends on the sample period and is not measured perfectly.

TABLE 9-1

Risk Models: Daily Data, 1990–1999

Parameter	Currency $/BP	DM/$	Yen/$	DM/BP	U.S. Stocks	U.S. Bonds	Crude Oil
Average SD σ(% pa)	11.33	10.54	11.78	7.98	14.10	4.07	37.55
GARCH process:							
α_0	0.00299	0.00576	0.01040	0.00834	0.00492	0.00138	0.04153
α_1	0.0379	0.0390	0.0528	0.1019	0.0485	0.0257	0.08348
β	0.9529	0.9476	0.9284	0.8699	0.9459	0.9532	0.9131
Persistence $(\alpha_1+\beta)$	0.9908	0.9866	0.9812	0.9718	0.9944	0.9789	0.9966

[4] As the number of observations T increases, the estimator converges to the true value.

FIGURE 9-4

GARCH volatility forecast.

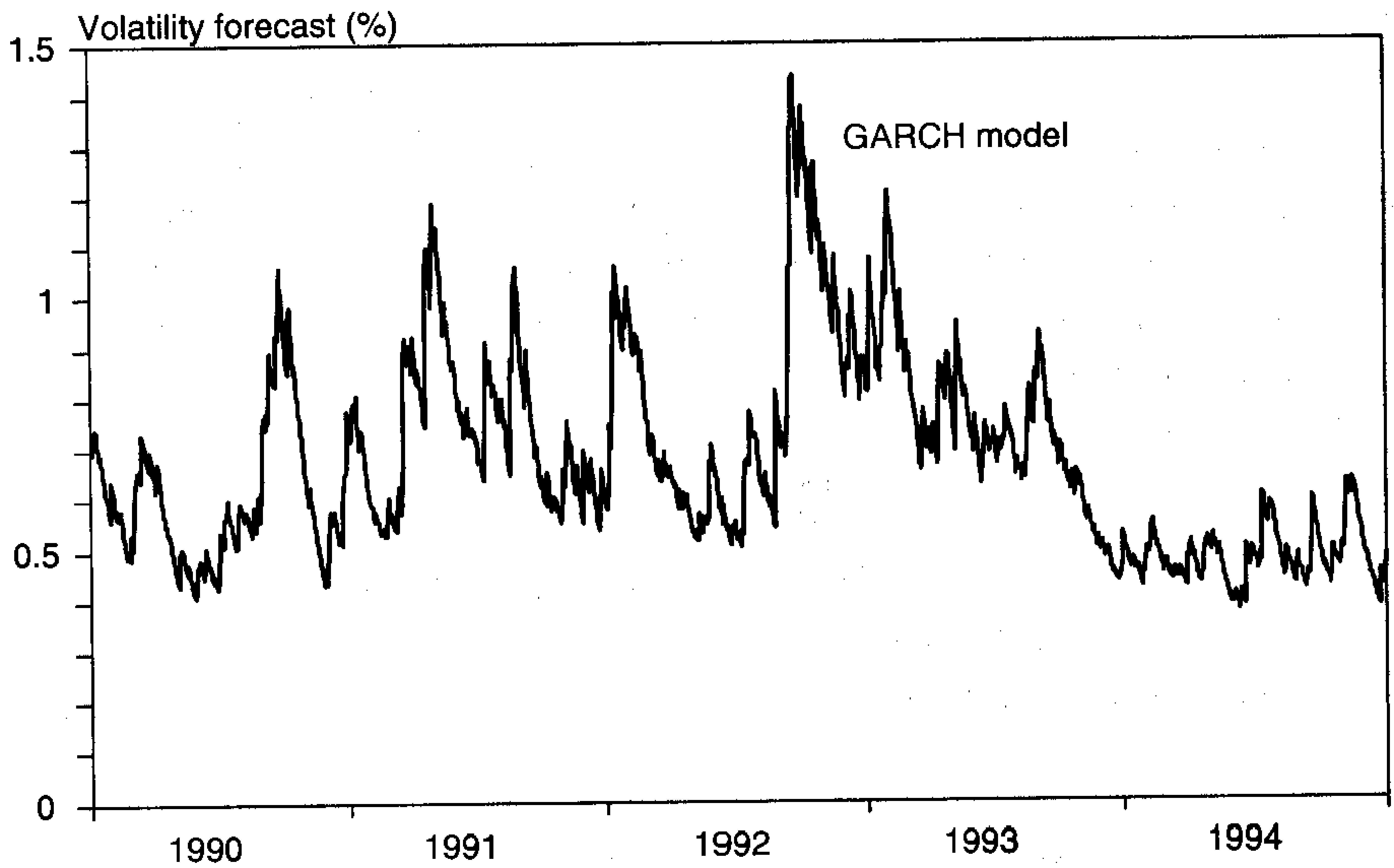


Figure 9-4 displays the GARCH forecast of volatility for the \$/BP rate. It shows increased volatility in the fall of 1992. Afterward, volatility decreases progressively over time, not in the abrupt fashion observed in Figure 9-3.

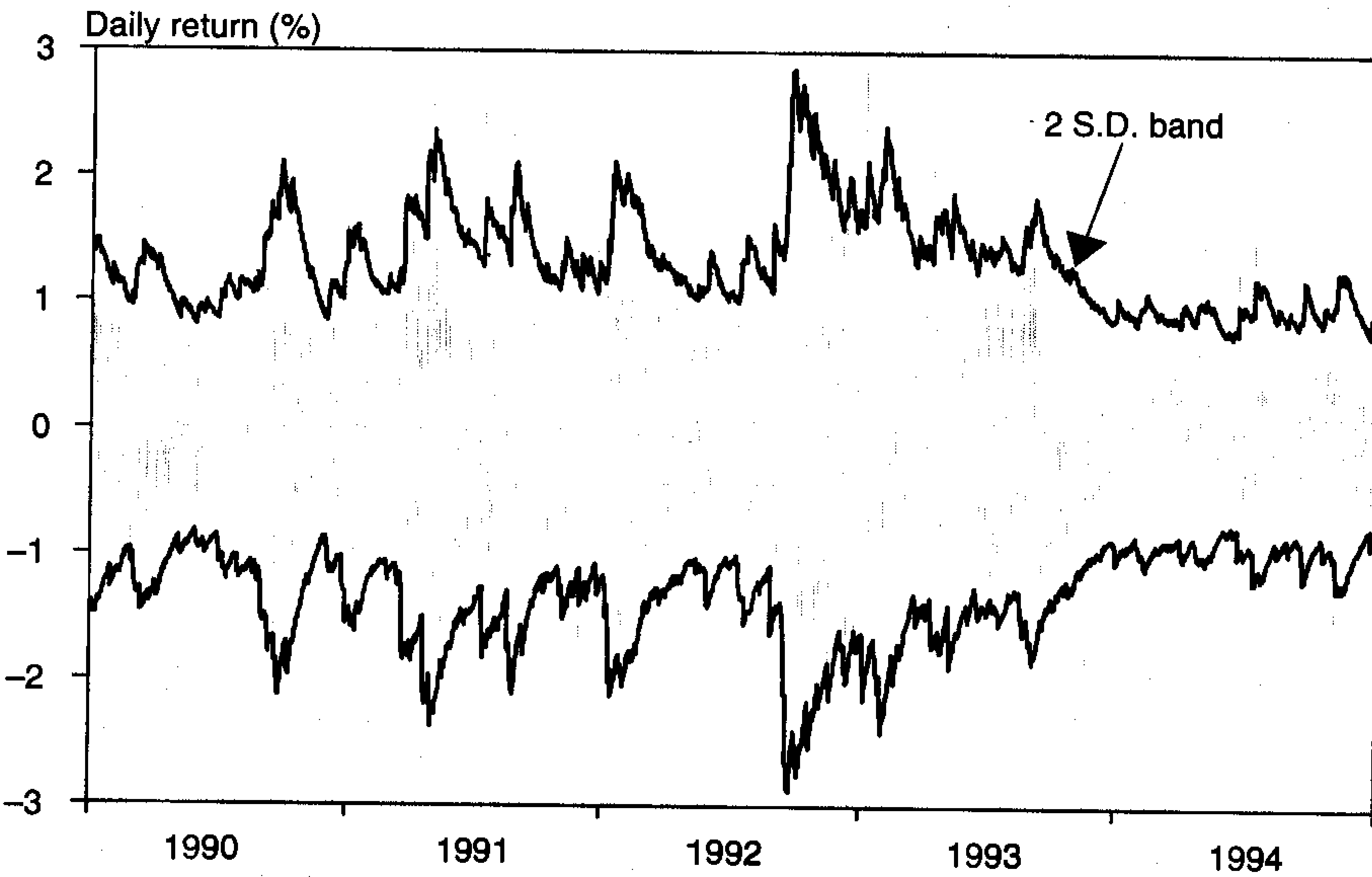
The practical use of this information is illustrated in Figure 9-5, which shows daily returns along with conditional 95 percent confidence bands, two-tailed, which involve plus or minus two standard deviations when the conditional residuals are normal. This model appears to capture variation in risk adequately. Most of the returns fall within the 95 percent band. The few outside the bands correspond to the remaining 5 percent of occurrences.

In practice, this basic GARCH model can be extended to other specifications. Because the innovation enters as a quadratic term, a day of exceptionally large value will have a very large effect on the conditional variance. This effect could be reduced by using the absolute value of the innovation instead.

Also, the basic GARCH model is symmetric. For some series, such as stocks, large negative returns have a bigger effect on risk than do positive

FIGURE 9-5

Returns and GARCH confidence bands.



returns, possibly reflecting a leverage effect.[5] GARCH models can be adapted to this empirical observation by using two terms, one for positive shocks and the other for negative shocks, each with its separate α coefficient.

9.2.3 Long-Horizon Forecasts

The GARCH model can be used to extrapolate the volatility over various horizons in a consistent fashion. Assume that the model is estimated using daily intervals. We first decompose the multiperiod return into daily returns as in Equation (4.27), that is,

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

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

[5] This asymmetry can be understood from the nonlinearity between the value of equity and the value of the firm. With fixed liabilities, a falling stock price increases the probability of bankruptcy and the risk of the stock. A stock-price appreciation decreases the probability of bankruptcy but by a lesser amount.

If returns are uncorrelated across days, the long-horizon variance as of $t - 1$ is

$$E_{t-1}(r_{t,T}^2) = E_{t-1}(r_t^2) + E_{t-1}(r_{t+1}^2) + E_{t-1}(r_{t+2}^2) + \cdots + E_{t-1}(r_T^2)$$

To determine the GARCH forecast in 2 days, we use tomorrow's forecast, that is,

$$E_{t-1}(r_{t+1}^2) = E_{t-1}(\alpha_0 + \alpha_1 r_t^2 + \beta h_t) = \alpha_0 + \alpha_1 h_t + \beta h_t$$

because $E_{t-1}(r_t^2) = h_t$. For the next day,

$$E_{t-1}(r_{t+2}^2) = E_{t-1}(\alpha_0 + \alpha_1 r_{t+1}^2 + \beta h_{t+1}) = \alpha_0 + (\alpha_1 + \beta)[\alpha_0 + (\alpha_1 + \beta) h_t]$$

Substituting n days into the future, the forecast of the "forward" variance at T is

$$E_{t-1}(r_T^2) = \alpha_0 \frac{1-(\alpha_1+\beta)^{n-1}}{1-(\alpha_1+\beta)} + (\alpha_1+\beta)^{n-1} h_t \tag{9.6}$$

The total variance from now to T then is

$$E_{t-1}(r_{t,T}^2) = \frac{\alpha_0}{1-(\alpha_1+\beta)}\left[(n-1) - (\alpha_1+\beta)\frac{1-(\alpha_1+\beta)^{n-1}}{1-(\alpha_1+\beta)}\right] + \frac{1-(\alpha_1+\beta)^n}{1-(\alpha_1+\beta)} h_t \tag{9.7}$$

This shows that the extrapolation of the next day's variance to a longer horizon is a complicated function of the variance process and the initial condition. Thus our simple square-root-of-time rule fails owing to the fact that returns are not identically distributed.

It is interesting to note that if we start from a position that is the long-run average, that is, $h_t = h = \alpha_0/[1-(\alpha_1+\beta)]$, this expression simplifies to

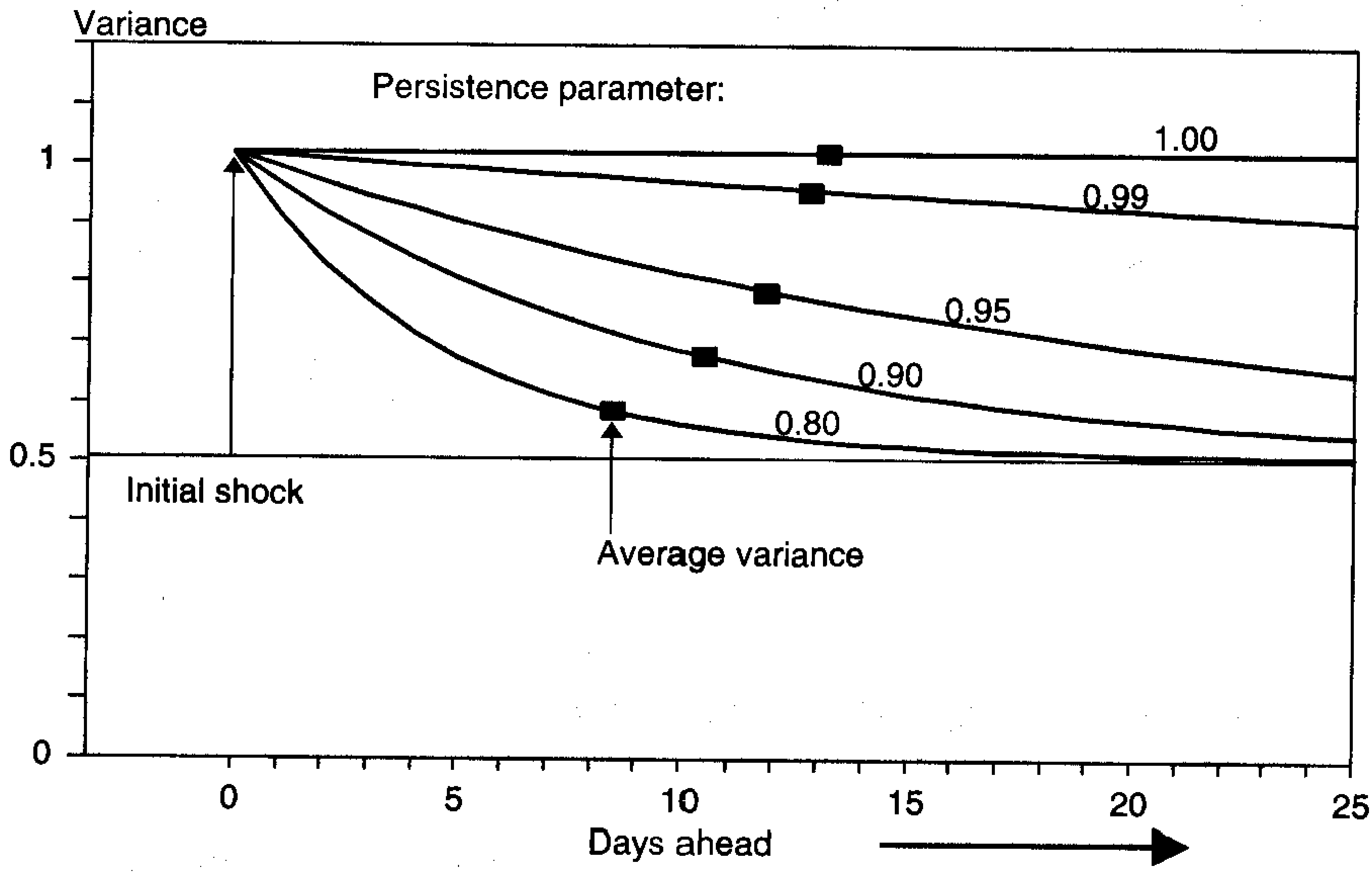
$$E_{t-1}(r_{t,T}^2) = hn \tag{9.8}$$

Here, the n-day volatility is the 1-day volatility times the square root of n. In other words, the extrapolation of VAR using the square root of time is only valid when the initial position happens to be equal to the long-run value. If the starting position is greater than the long-run value, the square-root-of-time rule will overestimate risk. If the starting position is less than the long-run value, the square-root-of-time rule will underestimate risk.

Figure 9-6 displays the effect of different persistence parameters $(\alpha_1 + \beta)$ on the variance. We start from the long-run value for the variance, that is, 0.51. Then a shock moves the conditional variance to twice

FIGURE 9-6

Mean reversion for the variance.



its value, about 1.02. This represents a very large shock. High persistence means that the shock will decay slowly. For instance, with persistence of 0.99, the conditional variance is still 0.93 after 20 days. With a persistence of 0.8, the variance drops very close to its long-run value after 20 days only. The marker on each line represents the average daily variance over the next 25 days.

Typical financial series have GARCH persistence of around 0.95 to 0.99 for daily data. In this situation, the figure shows that shocks decay quickly over long horizons, beyond 1 month. In fact, we could reestimate a GARCH process sampled at monthly intervals, and the coefficients α_1 and β would be much lower.[6] As a result, if the risk horizon is long, the swings in VAR should be much smaller than for a daily horizon. Christoffersen and Diebold (2000) even argue that there is scant evidence of volatility predictability at horizons longer than 10 days. Thus there is little point in forecasting time variation in volatility over longer horizons.

[6] Drost and Nijman (1993) provide closed-form solutions for this *temporal aggregation* issue. They show that the persistence decreases with longer horizons.

9.2.4 The RiskMetrics Approach

RiskMetrics takes a pragmatic approach to modeling risk.[7] Variances are modeled using an *exponentially weighted moving average* (EWMA) *forecast*. Formally, the forecast for time t is a weighted average of the previous forecast, using weight λ, and of the latest squared innovation, using weight $(1-\lambda)$, that is,

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

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

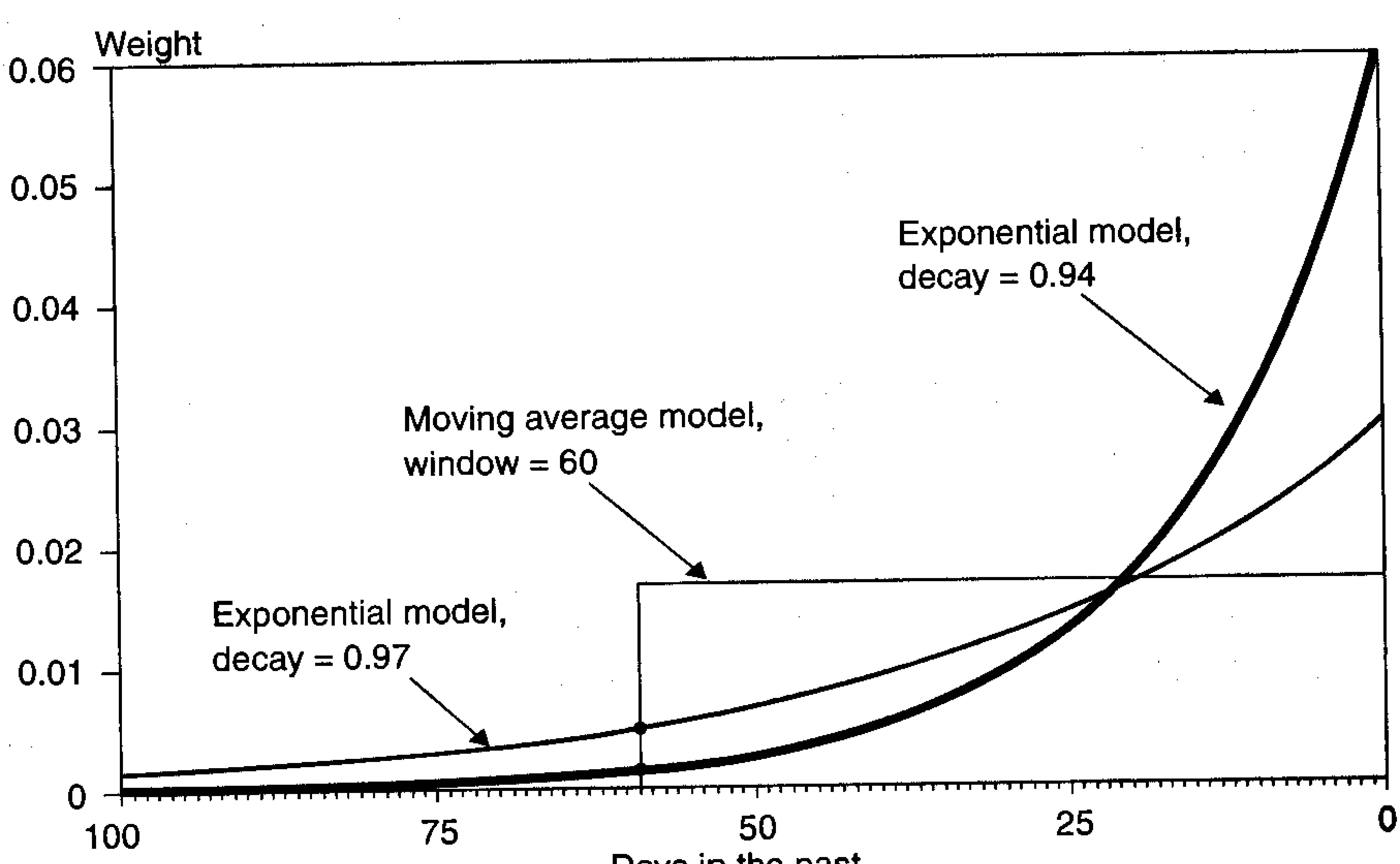
The exponential model places geometrically declining weights on past observations, thus assigning greater importance to recent observations. By recursively replacing h_{t-1} in Equation (9.9), we can write

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

Figure 9-7 displays the pattern of weights for $\lambda = 0.94$ and $\lambda = 0.97$. For $\lambda = 0.94$, the most recent weight is $1 - 0.94 = 0.06$. After that, the

FIGURE 9-7

Weights on past observations.



[7] For more detail on the methodology, see the RiskMetrics *Technical Manual* (1995).

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

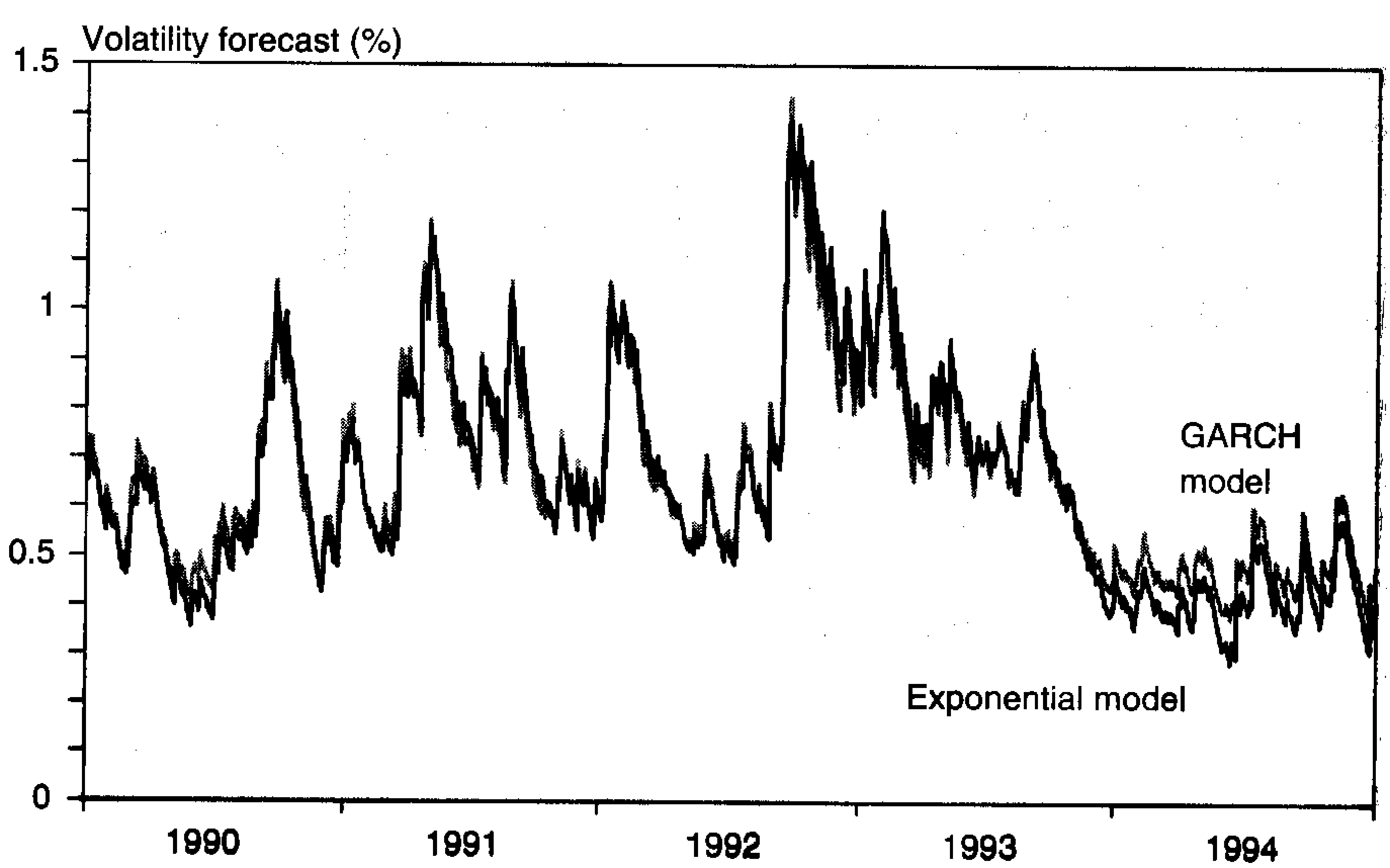
This model is a special case of the GARCH process where α_0 is set to 0 and α_1 and β sum to unity. The model therefore has persistence of 1. It is called *integrated GARCH* (IGARCH). As shown in Figure 9-8, the 1-day forecasts are nearly identical to those obtained with the GARCH model in Figure 9-4. 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.

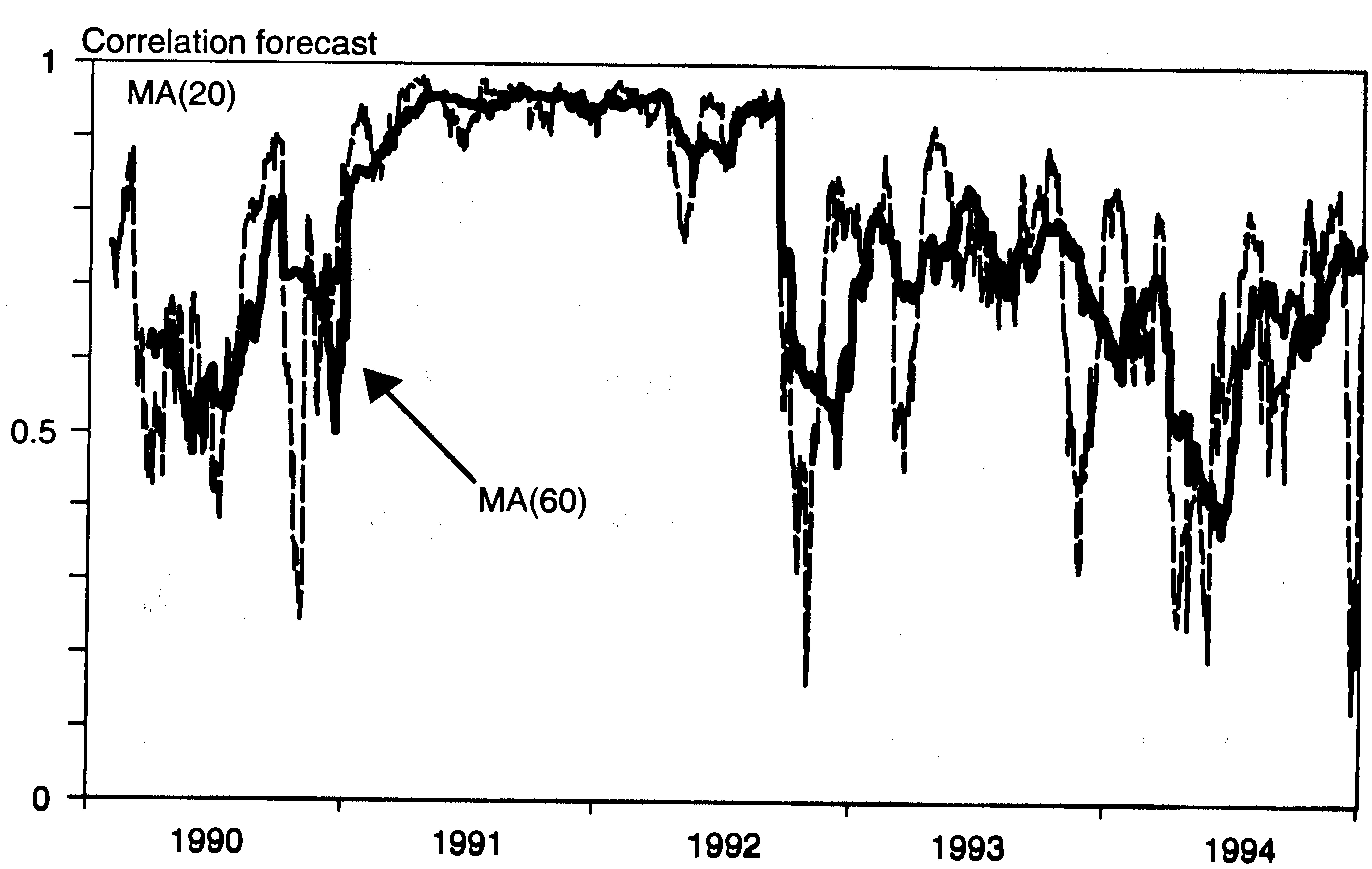
One advantage of multivariate volatility models is that they provide internally consistent risk estimates for a portfolio of assets. Another approach would be to construct the portfolio return series for given weights and to fit a univariate GARCH model to this aggregate series. If the weights change,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

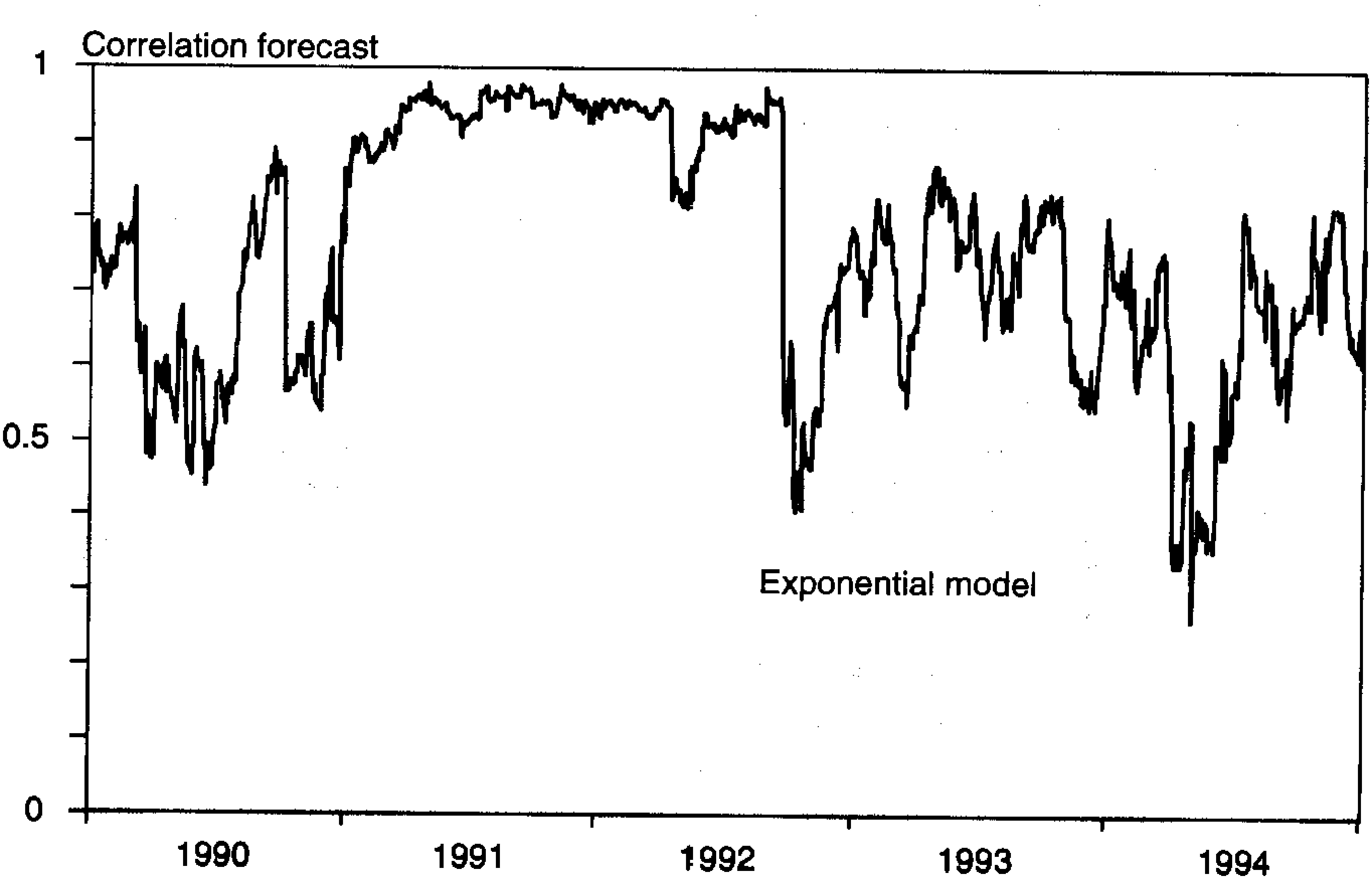
Note that the reason why RiskMetrics sets a common factor λ across all series is to ensure that all estimates of ρ are between -1 and 1. Otherwise, there is no guarantee that this will always be the case.

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,

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

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.

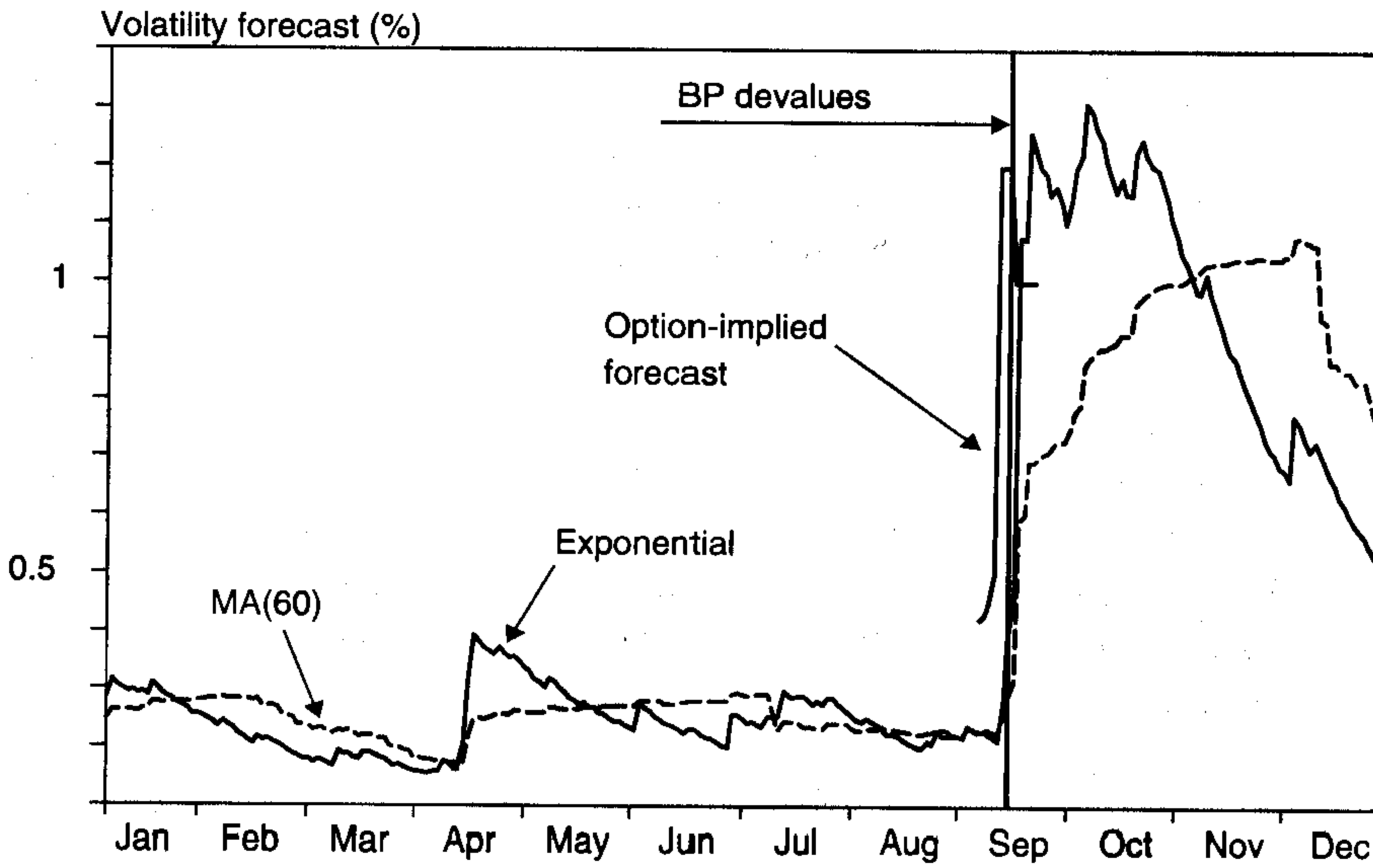
The empirical evidence indeed points to the superiority of options data.[9] An intuitive way to demonstrate the usefulness of options data is to analyze the September 1992 breakdown of the EMS. Figure 9-11 compares volatility forecasts during 1992, including that implied from DM/BP cross-options, the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