TABLE 8-6

Principal Components of Correlation Matrix: U.S. Bonds

	Eigenvectors			Percentage of Variance Explained by			
Maturity (year)	Factor 1 β_1	Factor 2 β_2	Factor 3 β_3	Factor 1	Factor 2	Factor 3	Total Variance Explained
1	0.27	0.52	0.79	72.2	17.9	9.8	99.8
2	0.30	0.34	−0.17	89.7	7.8	0.5	98.0
3	0.31	0.26	−0.22	94.3	4.5	0.7	99.5
4	0.31	0.18	−0.26	96.5	2.2	1.0	99.7
5	0.31	0.13	−0.24	97.7	1.1	0.9	99.7
7	0.31	−0.01	−0.17	98.9	0.0	0.4	99.3
9	0.31	−0.10	−0.11	98.2	0.7	0.2	99.1
10	0.31	−0.13	−0.08	98.1	1.2	0.1	99.4
15	0.30	−0.28	0.11	94.1	5.3	0.2	99.6
20	0.29	−0.41	0.24	87.2	11.0	0.9	99.1
30	0.29	−0.47	0.24	83.6	14.5	0.9	99.0
Average	0.30	0.00	0.01	91.9	6.0	1.4	99.3
Eigenvalue	10.104	0.662	0.156				

Table 8-6 displays the results of the PCA applied to the correlation matrix in Table 8-5.[3] Appendix 8.A gives more detail on the method.[4] With PCA, the factors are linear combinations of the data. Redefining the yield change dy as R, to shorten notations, the first principal component is defined as

$$z_1 = \beta_{11}R_1 + \cdots + \beta_{N1}R_N = \beta_1'R \tag{8.16}$$

Here, β_1 is called the first *eigenvector*, which represents the coefficients in the linear combination of the original variables that make up the first principal component. It is scaled so that the sum of its squared elements is 1. We observe from Table 8-6 that the first factor has similar coefficients across maturities. Thus it can be defined as a yield *level* factor.

[3] Note that PCA is sensitive to the values of the variances, or diagonal coefficients in the covariance matrix. Here we apply PCA to the correlation matrix. Using the correlation matrix gives equal weight to all risk factors.

[4] For the first application of PCA to the bond market, see Garbade (1986). Golub and Tilman (2000) provide a good overview of risk management with PCA.

TABLE 8-7

Correlation Matrix Fitted by First Component

Term (year)	1Y	2Y	3Y	4Y	5Y	7Y	9Y	10Y	15Y	20Y	30Y
1	0.722										
2	0.805	0.897									
3	0.825	0.920	0.943								
4	0.835	0.931	0.954	0.965							
5	0.840	0.936	0.959	0.971	0.977						
7	0.845	0.942	0.965	0.977	0.983	0.989					
9	0.842	0.939	0.962	0.974	0.979	0.985	0.982				
10	0.842	0.938	0.962	0.973	0.979	0.985	0.981	0.981			
15	0.824	0.919	0.942	0.953	0.959	0.965	0.961	0.961	0.941		
20	0.793	0.884	0.906	0.917	0.923	0.928	0.925	0.925	0.906	0.872	
30	0.777	0.866	0.888	0.898	0.904	0.909	0.906	0.906	0.887	0.854	0.836

The bottom of the table shows the associated *eigenvalues*, defined as the variance of z_1. For the first factor, this is $\sigma^2(z_1) = 10.104$.

Table 8-7 displays the correlation matrix fitted by the first principal component. This is constructed as $\beta_1\beta_1'\sigma^2(z_1)$. This matrix reproduces fairly well the large off-diagonal entries. Note that this matrix is very much simplified. In particular, it cannot be a true correlation matrix because the diagonal elements are not unity.

Going back to Table 8-6, the percentage of variance explained represents the fraction of the diagonal element explained by each principal component. For instance, this is 72.2 percent for the first risk factor and first maturity. This is also the first diagonal element in the fitted correlation matrix. Across all maturities, the average is 91.9 percent. Thus the first factor has high average explanatory power.

In economic terms, the *level* factor provides an excellent fit to movements of the term structure. This also explains why the *duration model* provides a good measure of interest-rate risk. The PCA approach, however, is slightly more general than duration because duration assumes first that all elements of the first eigenvector are identical and second that all yield volatilities are equal.

The second factor explains an additional 6.0 percent of movements. Because it has the highest explanatory power and highest loadings for short and long maturities, it describes the *slope* of the term structure. Finally, the last factor is much less important. It seems to be most related to 1-year rates, perhaps because of different characteristics of money-market instruments. Together, these three factors explain an impressive 99.3 percent of all return variation.

We now illustrate how PCA can be used to compute risk for sample portfolios. Because we use the correlation matrix of changes in yields, we need to convert these positions into dollar exposures on these normalized risk factors. Define these dollar exposures as x. Using Equation (8.14), each entry is

$$x = D^* \times \sigma(dy) \times P = D^* \times [\text{VAR}(dy)/1.65] \times \text{P} \qquad (8.17)$$

defining P as the market value of the position on each risk factor. The portfolio variance is then given by $x'\Sigma_\rho x$, where Σ_ρ is the correlation matrix of changes in yields.

With two factors, the portfolio variance is, from Equation (8.12),

$$\sigma^2(R_p) = \beta_{1p}^2\, \sigma^2\,(z_1) + \beta_{2p}^2 \sigma^2(z_2) \qquad (8.18)$$

where $\beta_{1p} = x'\beta_1$ is the portfolio exposure to the first factor, $\beta_{2p} = x'\beta_2$, to the second factor.

Consider first a portfolio investing P=\$100 million each in 1-year and 30-year bonds. As shown in Table 8-8, the first-factor exposure is $\beta_{1p} = 0.285 \times 0.2673 + 6.747 \times 0.2877 = 2.017$. The second is $\beta_{2p} = -3.005$. The portfolio variance is

$$\sigma^2(R_p) = (2.017^2 \times 10.104) + (-3.005^2 \times 0.662) = 41.099 + 5.977 = 47.076$$

Taking the square root and multiplying by $\alpha = 1.65$, this gives a two-factor portfolio VAR of \$11.32 million. Using the full 11 factors gives a VAR of \$11.44 million. The first factor alone would have given a VAR of $\sqrt{111.892}$ = \$10.58 million, which is close. Thus, for this simple portfolio, using one principal component only would provide a good approximation to the true risk.

Table 8-8 analyzes another portfolio with \$100 million invested in the 10-year bond, \$40 million short the 30-year bond, and \$60 million short the 1-year bond. Because of the long and short positions, this is

TABLE 8-8

Risk Analysis by Principal Components

Maturity (year)	Modified Duration D^*	Yield VAR (%) VAR(dy)	Portfolio 1 Position ($ million) P	x	Portfolio 2 Position ($ million) P	x
1	0.945	0.497	+100	0.285	−60	−0.171
2	1.892	0.522	0	0	0	0
3	2.835	0.523	0	0	0	0
4	3.777	0.522	0	0	0	0
5	4.719	0.514	0	0	0	0
7	6.599	0.484	0	0	0	0
9	8.475	0.462	0	0	0	0
10	9.411	0.452	0	0	+100	2.578
15	14.072	0.443	0	0	0	0
20	18.737	0.435	0	0	0	0
30	28.111	0.396	+100	6.747	−40	−2.699
Exposure				$\beta_{1p} = +2.017$		$\beta_{1p} = -0.019$
				$\beta_{2p} = -3.005$		$\beta_{2p} = +0.829$

largely hedged against the first factor, with $\beta_{1p} = -0.019$ only. The two-factor risk analysis gives

$$\sigma^2(R_p) = (-0.019^2 \times 10.104) + (0.829^2 \times 0.662) = 0.004 + 0.455 = 0.459$$

Using the one-factor model generates a VAR of \$0.10 million, which is too low. The two-factor model provides a better approximation, a VAR of \$1.12 million that is close to the true VAR of \$1.42 million. Here we need at least a two-factor model.

This decomposition shows that for some purposes, the risk of a bond portfolio can be usefully summarized by its exposure to a very small number of factors. Whether this is sufficient depends on the structure of the portfolio being modeled.[5]

[5] As an example, Jamshidian and Zhu (1997) use information from PCA to design simulations that are very efficient. The basic idea is to simulate more data points for the first factor and fewer for remaining factors. Gibson and Prisker (2001), however, show that this approximation may fail, especially for hedged portfolios, and propose refinements to correct these shortcomings.

BOX 8-1

RISK MODELS AT PIMCO

The *Total Return Fund*, run by asset manager Pacific Investment Management Company (PIMCO), is the largest bond mutual fund in the world, with close to $100 billion in assets. The portfolio has more than 10,000 different positions in fixed-income instruments, including derivatives. It would be impossible for the portfolio manager, Bill Gross, to keep track mentally of all these positions. This is where risk models can help.

PIMCO reduces the dimensionality of the problem by focusing on a small number of risk factors. These include (1) the level of the yield curve, (2) the slope between the 2- and 10-year maturities on the yield curve, (3) the slope between the 10- and 30-year maturities, (4) the spread between mortgages and Treasuries, and (5) the spread between corporates and Treasuries. Each position is expressed in terms of its exposure to these risk factors. These exposures are totted up across the entire portfolio, giving summary measures of exposures to these principal risk factors. The portfolio manager then can translate bets on risk factors into exposures and positions.

In 2002, PIMCO received the *asset management risk manager of the year award*. In describing this prestigious award, *Risk* noted that the "firm's risk-centric decision-making has allowed it to consistently beat its benchmarks." Indeed, over the previous 10 years, the Total Return Fund has rewarded investors with an average value added of 1.5 percent annually.

8.2.5 Comparison of Methods

To illustrate this important point, Table 8-9 presents VAR calculations for three portfolios.[6] The first is a diversified portfolio with $1 million equally invested in 10 stocks. The second consists of a $1 million portfolio with 10 stocks all in the same industry (high technology). The third expands on the diversified portfolio but is market-neutral, with long positions in the first five stocks and short the others. In other words, this is a *hedge fund* with zero net position in stocks.

[6] The diversified portfolio consists of positions in Ford, Hewlett-Packard, General Electric, Procter & Gamble, AT&T, Boeing, General Motors, Disney, Microsoft, and American Express. These are spread among 6 of the 10 industrial sectors in the market. The long-short portfolio is long the first five and short the others. The market index is taken as the Standard & Poor's (S&P) 500. VAR is measured with a 1-month horizon at the 95 percent level of confidence using historical data from 1990 to 1999.

TABLE 8-9

Comparison of VAR Methods

	Portfolio		
	Diversified	High Tech	Long-Short
Net Position	$1,000,000	$1,000,000	$0
VAR			
Index mapping	$63,634	$63,634	$0
Beta mapping	$70,086	$84,008	$298
Industry mapping	$69,504	$90,374	$7,388
Diagonal model	$81,238	$105,283	$41,081
Individual mapping (exact)	$78,994	$118,955	$32,598

Five methods are examined:

- *Index mapping* replaces each stock by a like position in the index m, that is,

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

- *Beta mapping* only considers the net beta of the portfolio, that is,

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

- *Diagonal model* considers both the beta and specific risk, that is,

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

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

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

- *Individual mapping* uses the full covariance matrix of individual stocks and provides an exact VAR measure over this sample period, that is,

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

The table shows that the quality of the approximation depends on the structure of the portfolio. This is an important conclusion. For the first portfolio, all measures are in a similar range, \$60,000–\$80,000. The diagonal model provides the best approximation, followed by the beta and industry-mapping models.

The second portfolio is concentrated in one industry and, as a result, has higher VAR. The index-mapping model now seriously underestimates the true risk of the portfolio. In addition, the beta and industry-mapping models also fall short because they ignore the portfolio concentration. The diagonal model is closest to the exact value, as before.

Finally, the third portfolio shows the dangers of simple mapping methods. The index-mapping model, given a zero net investment in stocks, predicts zero risk. With beta mapping, the risk measure, driven by the net beta, is close to zero, which is highly misleading. The best approximation is again provided by the diagonal model, which considers specific risks. In conclusion, the best risk model depends on the portfolio. This requires risk managers to have a thorough understanding of the investment process.

8.3 COPULAS

The traditional approach to multivariate analysis is based on the joint multivariate normal distribution for the risk factors. This implies that expected returns are linearly related to each other, as described by correlation coefficients and that, in addition, the probability of seeing extreme observations for many risk factors is low. A growing body of empirical research, however, indicates that these assumptions may be suspect. And this matters: The joint tail behavior of risk factors drives the shape of the tails of the portfolio distribution. Thus, using a normal assumption could lead to a serious underestimation of value at risk.

8.3.1 What Is a Copula?

This is where the concept of copulas comes to the rescue. To simplify, consider two risk factors only, 1 and 2. Their joint distribution can be split up into two statistical constructs. First is the marginal distribution for the two variables, $f_1(x_1)$ and $f_2(x_2)$. Second is the way in which the two marginals are *attached* to each other. This is done with a *copula*, which is a function that links marginal distributions into a joint distribution. Formally, the copula is a function of the marginal (cumulative) distributions $F(x)$, which range from 0 to 1. In the bivariate case, it has two arguments plus parameters θ, that is,

$$c_{12}\ [F_1(x_1), F_2(x_2);\theta] \tag{8.19}$$

The link between the joint and marginal distributions is made explicit by *Sklar's theorem*, which states that for any joint density there exists a copula that links the marginal densities, that is,

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2);\theta] \tag{8.20}$$

Consider, for example, a multivariate *normal* distribution. This can be split into two normal marginals and a normal copula. Assume that all variables are standardized, that is, have zero mean and unit standard deviation. Define Φ as the normal probability density function, N as the cumulative normal function, c^N as the normal copula, and ρ as its correlation coefficient. This gives

$$f_1(x_1) = \Phi(x_1) \qquad f_2(x_2) = \Phi(x_2) \tag{8.21}$$

and

$$f_{12}(x_1,x_2) = \Phi(x_1,x_2;\rho) = \Phi(x_1) \times \Phi(x_2) \times c_{12}^N\ [N(x_1), N(x_2);\rho] \tag{8.22}$$

This shows that a bivariate normal density is constructed from two normal marginal densities and a normal copula. The bivariate density has one parameter, the correlation coefficient, which only appears in the copula.

Thus the copula contains all the information on the nature of the dependence between the random variables but gives no information on the marginal distributions. This allows a neat separation between the marginals and dependence. More complex dependencies can be modeled with different copulas.

8.3.2 Marginals and Copulas

In general, the copula can be any function that satisfies the appropriate restrictions behind Equation (8.20). It can be derived from the joint density function, for example, the normal or the student t. The student distribution is interesting because it displays fatter tails than the normal and greater dependences in the tails. We could mix and match the normal and student marginals with the normal and student copulas to represent the data better.

FIGURE 8-2

Combination of marginals and copulas.

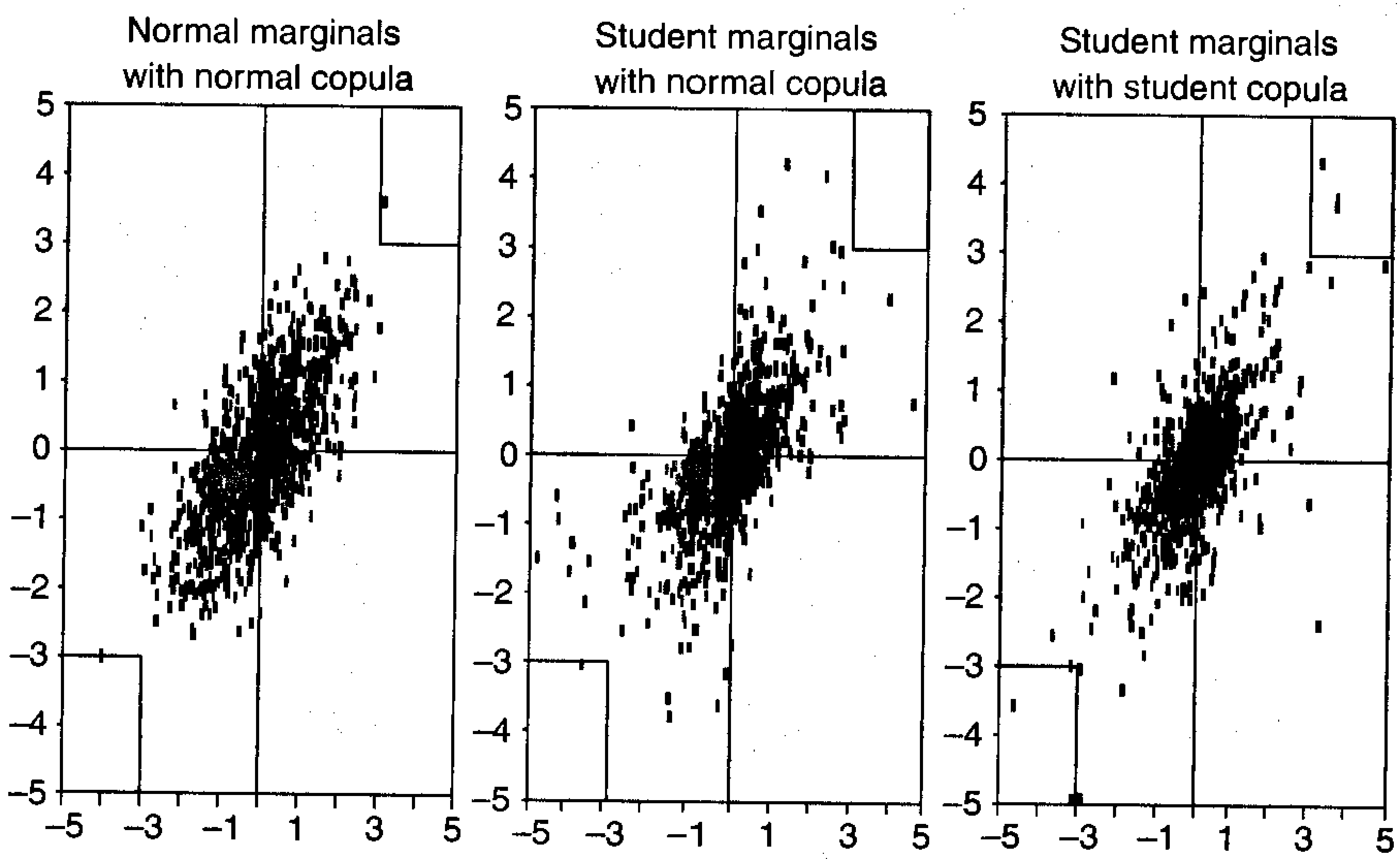


Figure 8-2 describes a plot of two variables generated with (1) normal marginals with a normal copula, (2) student marginals with a normal copula, and (3) student marginals with a student copula parametrized with 3 degrees of freedom. Now observe the boxes including extreme observations, either both above +3 or both below −3. If the two series are returns on two stock markets, for instance, we should be worried about situations where the two markets fall at the same time because this increases the risk of portfolios that have long positions in the two markets.

With the normal marginals, the dispersion of each variable is limited. Indeed, the probability of a move beyond +3 or −3 is only 0.003 percent or, on average, 3 observations from the 1000 in this sample. With the student marginals, there is greater dispersion of each variable, reflecting the fatter tails of the distribution. The two left panels in Figure 8-2 are based on the normal copula. This does not generate many dependencies in the tails. In the panel on the left we only have two cases with joint extreme observations; in the middle panel, there is only one.

The panel on the right combines the student marginals with the student copula. This has many more observations in the tails, three in the top box and three in the lower box. These comovements increase the portfolio risk sharply. As an example, consider a portfolio equally weighted in the two risk factors. The VAR at the 99 percent confidence level increases from 2.1 to 2.3 to 2.4 when going from the left to the right panel. In this case, assuming a normal distribution understates risk by more than 10 percent, which is substantial. Furthermore, this bias will worsen with a greater number of risk factors. If the student copula is a better reflection of reality, it should be used instead of the normal copula.

To summarize, the risk-modeling process works in several steps. First, the risk manager has to choose the best functional form for the marginal distributions and the copula function. Second, the parameters of these functions must be estimated. The third step then consists of running simulations that generate random variables that mimic the risk factors. The current portfolio can be modeled as a series of positions on the risk factors. In the final step, the risk manager constructs the distribution of returns for the current portfolio. This can be summarized by VAR using a quantile of the distribution.

8.3.3 Applications

The preceding section has illustrated the use of *elliptical* copulas, which are symmetric around the mean. These imply the same probability of joint positive or negative movements, assuming positive correlations.

More generally, the copula can be asymmetric, with greater probability of joint moves in one direction or another. Geman and Kharoubi (2003), for example, wanted to examine the association between stocks and hedge-fund strategies. They fit several copulas to the joint movements between historical series. They found that for most categories of hedge funds, the *Cook-Johnson copula* provides the best fit. This is an asymmetric copula with greater probability of joint down moves for the two risk factors. This means that when stock markets drop precipitously, it is likely that some hedge-fund strategies will lose money as well. As a result, some categories of hedge funds provide much less diversification with stocks than hoped for.

Copulas are bound to be used increasingly in financial risk management because they can be used to build joint distributions of risk factors. They are finding a wide range of applications, as illustrated in Box 8-2. Another application, detailed in Chapter 21, will be the integration of market, credit, and operational risk at the highest level of the financial institution.

BOX 8-2

COPULAS IN FINANCE

Collateralized debt obligations (CDOs) are pooled investments in debt instruments that offer ready-made diversification. The total cash flows are directed to different classes of claims, or *tranches*, according to predefined priority rules. Losses owing to default hit first the lowest-rated tranches, then the middle-rated tranches (called *mezzanine*), and then the senior tranches. To ascertain expected losses to each tranche, we need to construct the entire distribution of portfolio values.

Payoffs on CDO tranches depend heavily on correlations among defaults in the underlying credit portfolio. Low correlations make the senior tranches safer. On the other hand, if all underlying bonds default at the same time, the senior tranches could face serious losses. David Li (2000) is widely credited with having developed the first commercial model for CDO pricing, using the concept of copula functions.

Since then, the standard industry model has been the normal copula because of its simplicity. CreditMetrics, for instance, generates a joint distributions in asset values using a multivariate normal distribution, which implies a normal copula.

Like all models, these are just approximations of reality. Sometimes these approximations work poorly. On May 5, 2005, the credit-rating agencies downgraded the debt issued by General Motors (GM) and Ford to below investment grade. This event, however, was specific to these two firms and did not affect others. Many investors had tried to hedge GM and Ford's debt by shorting other bonds, based on the relationships predicted by the normal copula. They lost millions of dollars during this episode.

8.4 CONCLUSIONS

Risk management systems typically involve large-scale aggregation. Because of the number of risk factors, simplifications are often required. This chapter has provided tools to model the multivariate distribution of risk factors.

This involves choosing the shape of the joint density and its parameters. Generally, normal joint densities are used merely because of convenience. Such densities, however, do not generate the joint movements in the tails that we seem to observe in empirical data. This is important because the possibility of large simultaneous drops in prices means that the portfolio risk can be very high. Such tail dependences can be modeled, for instance, using the student copula.

The covariance matrix, or correlation matrix, also needs special attention. In large samples, portfolio risk is driven primarily by correlations. With a large number of assets, however, there are too many parameters to estimate. The covariance matrix needs simplifications. Factor models help to reduce the dimensionality of the problem.

A particularly interesting application is that of principal component analysis. This approach simplifies the risk measurement process considerably and gives a better understanding of the underlying economics. The choice of number of risk factors, however, is driven by a tradeoff between parsimony and accurate risk measurement. Ultimately, the choice of the joint distribution should be made by the risk manager based on market experience and a solid understanding of these multivariate models.

APPENDIX 8.A

Principal Component Analysis

Consider a set of N variables $R_1, \ldots, R_N$ with covariance matrix Σ. These could be bond returns or changes in bond yields, for instance. We wish to simplify or reduce the dimensions of Σ without too much loss of content by approximating it by another matrix Σ^*. Our goal is to provide a good approximation of the variance of a portfolio $R_p = w'R$ using $V^*(R_p) = w'\Sigma^* w$. The process consists of replacing the original variables R by another set z suitably selected.

The *first* principal component is the linear combination

$$z_1 = \beta_{11}R_1 + \cdots + \beta_{N1}R_N = \beta_1'R \tag{8.23}$$

such that its variance is maximized, subject to a normalization constraint on the norm of the factor exposure vector $\beta_1'\beta_1 = 1$. A constrained optimization of this variance, $\sigma^2(z_1) = \beta_1'\Sigma\beta_1$, shows that the vector β_1 must satisfy $\Sigma\beta_1 = \lambda_1\beta_1$. Here, $\sigma^2(z_1) = \lambda_1$ is the largest *eigenvalue* of the matrix Σ, and β_1 its associated *eigenvector*.

The *second* principal component is the one that has greatest variance subject to the same normalization constraint $\beta_2'\beta_2 = 1$ and to the fact that it must be orthogonal to the first $\beta_2'\beta_1 = 0$. And so on for all the others.

This process basically replaces the original set of R variables by another set of z orthogonal factors that has the same dimension but where the variables are sorted in order of decreasing importance. This leads to the *singular value decomposition*, which decomposes the original matrix as

$$\Sigma = PDP' = [\beta_1 \ldots \beta_N] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_N' \end{bmatrix} \tag{8.24}$$

where P is an orthogonal matrix, that is, such that its inverse is also its transpose, $P^{-1} = P'$, and D a diagonal matrix composed of the λ_i's. The next step would be to give an economic interpretation to the principal components by examining patterns in the eigenvectors.

The definition of P implies that we can write the transformation conveniently as $z = P'R$. Alternatively, if we are given the set of y, we can recover R as $R = Pz$. In other words,

$$R_i = \beta_{i1} z_1 + \cdots + \beta_{iN} z_N \tag{8.25}$$

To each z_j is associated a value for its variance λ_j that is sorted in order of decreasing importance. These eigenvalues are quite useful because they can tell us whether the original matrix Σ truly has N dimensions. For instance, if all the eigenvalues have the same size, then all transformed variables are equally important. In most situations, however, some eigenvalues will be very small, which means that the true dimensionality (or rank) is less than N.

In other cases, some values will be zero or even negative, which indicates that the matrix is not defined properly. The problem is that for some portfolios, the resulting VAR could be negative![7]

If so, we can decide to keep only the first K components, beyond which their variances λ_j can be viewed as too small and unimportant. Thus we replace the previous exact relationship by an approximation, that is,

$$R_i \approx \beta_{i1} z_1 + \ldots + \beta_{iK} z_K \tag{8.26}$$

Based on this, we approximate the matrix by

$$\Sigma^* = [\beta_1 \cdots \beta_K] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_K \end{bmatrix} \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_K' \end{bmatrix} = \beta_1 \beta_1' \lambda_1 + \cdots + \beta_K \beta_K' \lambda_K \tag{8.27}$$

which is very close to Equation (8.11), except for the residual terms on the diagonal. Note that this matrix Σ^* is surely not invertible because it has only rank of K by construction yet has dimension of N.

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

The benefit of this approach is that we can now simulate movements in the original variables by simulating movements with a much smaller set of variables z, called *principal components* (PCs). The fraction of the variance explained, as reported in Table 8-6, is given by the diagonal of this matrix. For the first asset, for instance, the first PC explains a fraction of $\beta_{11}^2\lambda_1/\sigma_1^2$, the second $\beta_{12}^2\lambda_2/\sigma_1^2$, and so on.

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

$$
\begin{aligned}
R_p &= \Sigma w_i R_i \approx w_1(\beta_{11}z_1 + \cdots + \beta_{1K}z_K) \cdots + w_N(\beta_{N1}z_1 + \cdots + \beta_{NK}z_K) \\
&= (w_1\beta_{11} + \cdots + w_N\beta_{N1})z_1 + \cdots + (w_1\beta_{1K} + \cdots + w_N\beta_{NK})z_K \\
&= \delta_1 z_1 + \cdots + \delta_K z_K
\end{aligned}
$$

Each term between parentheses represents the weighted exposure to each principal component. For instance, $\delta_1 = w'\beta_1$ would be the portfolio exposure to the first PC. In the stock market, this would be the portfolio total systematic risk. This decomposition is useful for performance attribution because it breaks down the portfolio return into the exposure and return on each PC.

In addition, we can compute the variance of the portfolio directly from Equation (8.27):

$$
\begin{aligned}
\sigma^2(R_p) &= w'\Sigma^* w = w'\beta_1\beta_1' w\lambda_1 + \cdots + w'\beta_K\beta_K' w\lambda_K \\
&= (w'\beta_1)^2\lambda_1 + \cdots + (w'\beta_K)^2\lambda_K \\
&= \delta_1^2\sigma^2(z_1) + \cdots + \delta_K^2\sigma^2(z_K)
\end{aligned}
\tag{8.28}
$$

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

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

QUESTIONS

1. What is the main drawback of the analytical approach to measure VAR based on the full covariance matrix with a large number of assets?
2. Give examples of situations where the covariance matrix is not positive definite.

3. Consider the following covariance matrix: $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
   (a) Is this positive definite?
   (b) What is the meaning of a unit correlation coefficient?
   (c) Can we come up with a vector of positions that will create zero risk?
4. Consider two stocks with the following decomposition on the market index:

	α	β	σ_ϵ
A	0.10	0.8	0.12
B	0.05	1.2	0.20

   The volatility of the market index is 0.15. Compute the covariance matrix using the diagonal and beta models. What is the correlation between the stocks?
5. The factor model behind the driver for asset correlations under Basel II is $R_i = \sqrt{0.2}R_m + \sqrt{1-0.2}\epsilon_i$, where the volatility of R_m and ϵ is 1; the residuals ϵ are uncorrelated across assets. Compute the correlation between any two assets.
6. From the principal components analysis of correlation matrix of U.S. bonds, how many primitive risk factors can represent movements in the yield curve?
7. The duration model is similar, but not identical, to using a first principal component only for fixed-income securities. What are the differences?
8. Using data from section 8.2.4, compute the one-factor and two-factor VAR measures for the portfolios: (a) \$100 million in the 5-year bonds and \$100 million in the 10-year bonds (b) short \$170 million in the 5-year bonds and long \$100 million in the 10-year bonds. Comment on the results.
9. In what situations will index mapping fail?
10. A risk manager wants to assess the risk of a hedge fund. The fund is concentrated in a few stocks and is market-neutral (in other words, it has zero beta). Under these conditions, is it appropriate to use a one-factor market model?
11. Factor analysis reduces correlations to interactions between a small number of risk factors. For fixed-income portfolios, the number of factors should be two, irrespective of the portfolio. Discuss.

12. A portfolio manager invests in the U.S. and euro bond markets. Returns are measured in dollars. How many important factors are likely to show up in a principal components analysis applied to these markets?
13. Copulas are functions that attach marginal densities to form joint densities. For the normal copula, should the mean and standard deviation of each marginal enter as parameters in the copula?
14. Which of the following three combinations should generate the highest probability of large joint losses? (a) A normal multivariate density, (b) a normal copula with student marginals, (c) a student multivariate density.
15. Can copulas be used to model nonlinear correlation coefficients?
16. Why is the shape of the copula important to assess the possibility of losses in senior CDO tranches?

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.

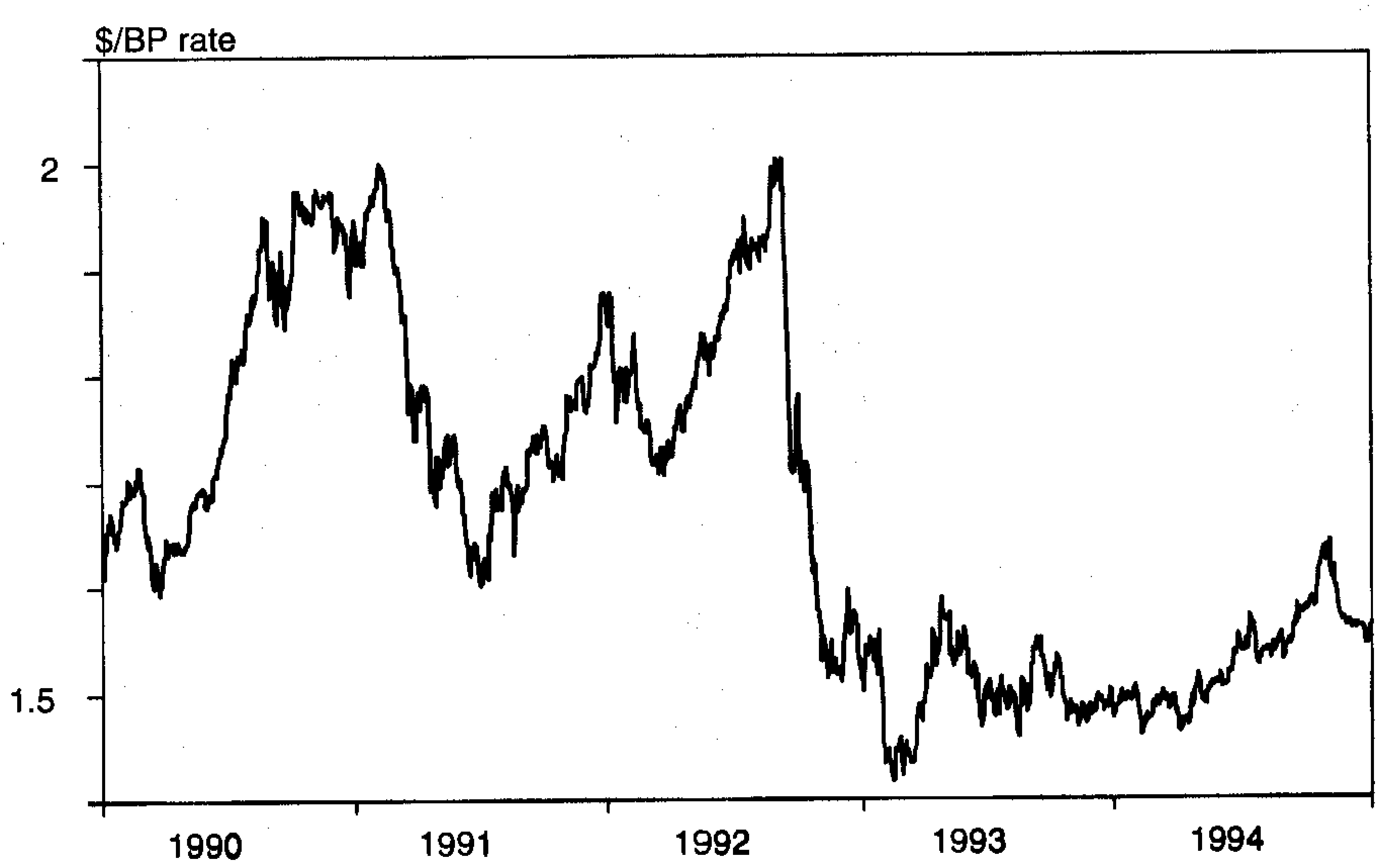
The purpose of this chapter is to present techniques to forecast variation in risk and correlations. Section 9.1 motivates the problem by taking the example of a series that underwent structural changes leading to predictable patterns in volatility. Section 9.2 then presents recent developments in time-series 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.

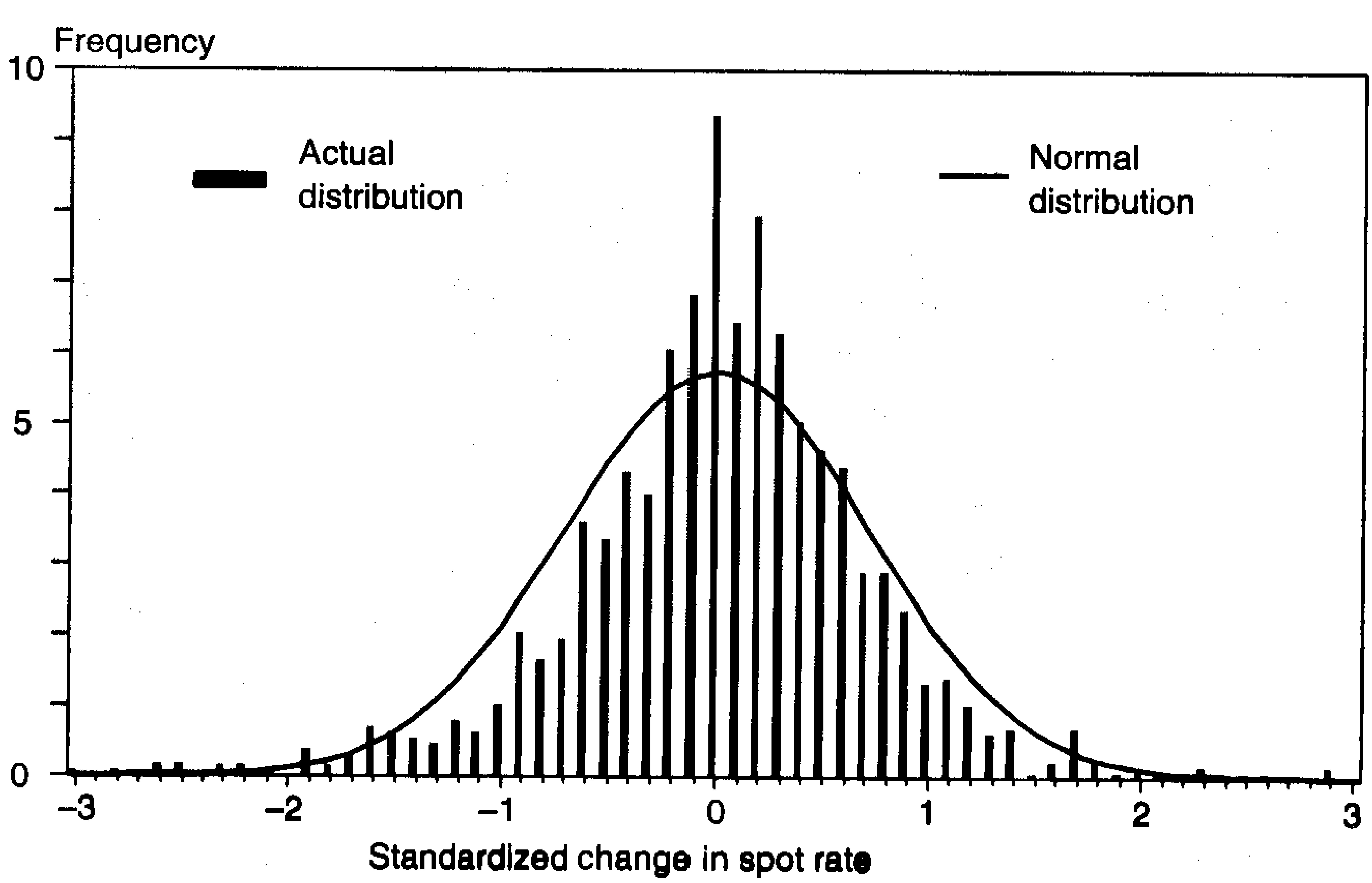
Figure 9-2 compares the normal approximation with the actual empirical distribution of the $/BP exchange rate. Relative to the normal model, the actual distribution contains more observations in the center and in the tails.

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.

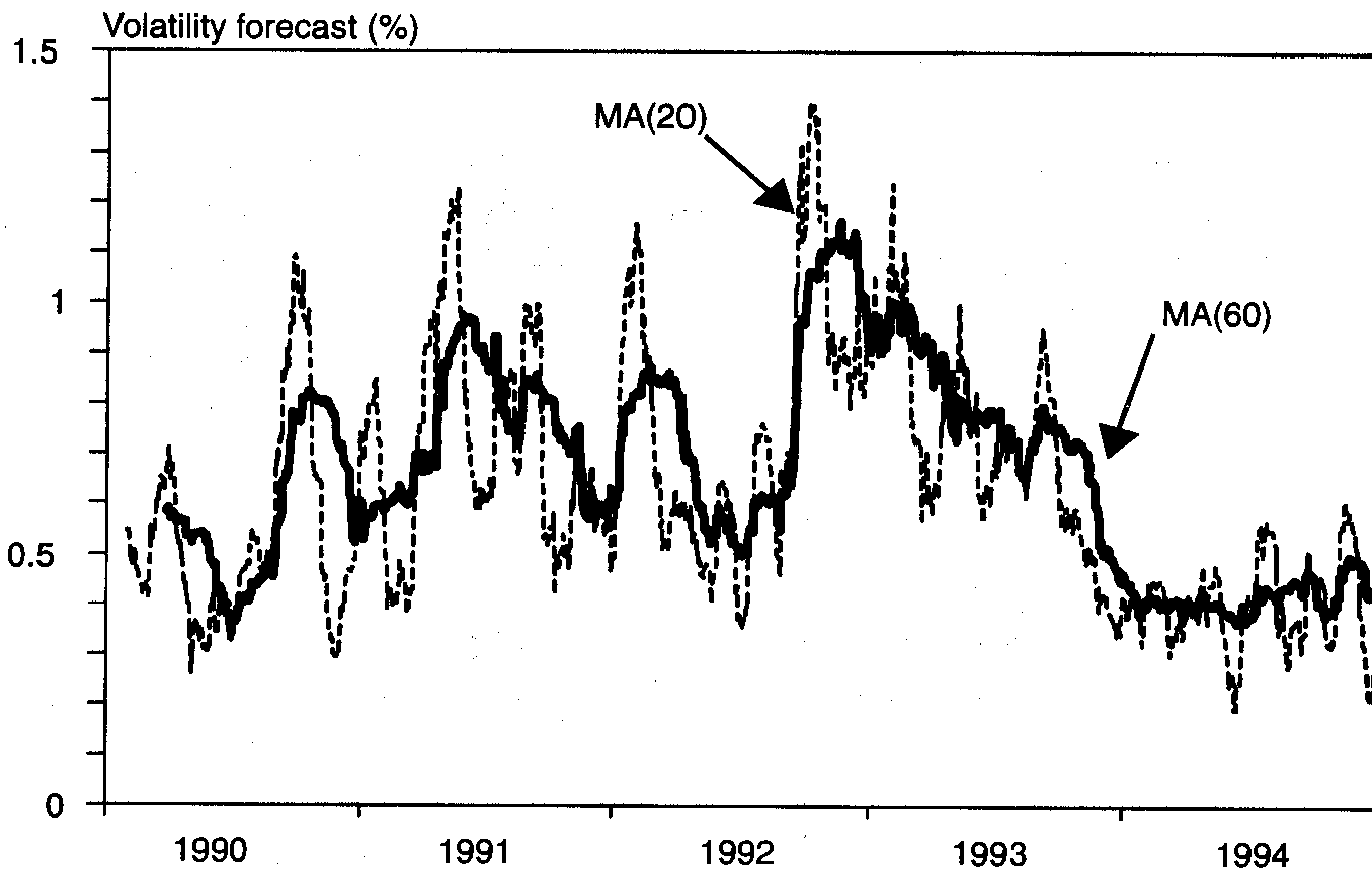
While simple to implement, this model has serious drawbacks. First, it ignores the dynamic ordering of observations. Recent information receives the same weight as older observations in the window that may no longer be relevant.

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,

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

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.