Hung, Shackleton and Xu, 2004). A recent study by Brooks, Černý and Miffre (2012) proposed a utility-based framework for the determination of optimal hedge ratios that can allow for the impact of higher moments on the hedging decision in the context of hedging commodity exposures with futures contracts.

9.19.2 Tail Models

It is widely known that financial asset returns do not follow a normal distribution, but rather they are almost always *leptokurtic*, or *fat-tailed*. This observation has several implications for econometric modelling. First, models and inference procedures are required that are robust to non-normal error distributions. Second, the riskiness of holding a particular security is probably no longer appropriately measured by its variance alone. In a risk management context, assuming normality when returns are fat-tailed will result in a systematic underestimation of the riskiness of the portfolio. Consequently, several approaches have been employed to systematically allow for the leptokurtosis in financial data, including the use of a Student's *t* distribution.

Arguably the simplest approach is the use of a mixture of normal distributions. It can be seen that a mixture of normal distributions with different variances will lead to an overall series that is leptokurtic. Second, a Student's *t* distribution can be used, with the usual degrees of freedom parameter estimated using maximum likelihood along with other parameters of the model. The degrees of freedom estimate will control the fatness of the tails fitted from the model. Other probability distributions can also be employed, such as the 'stable' distributions that fall under the general umbrella of extreme value theory – see Section 14.3 of Chapter 14 for a detailed presentation of this class of models.

9.20 Forecasting Covariances and Correlations

A major limitation of the volatility models examined above is that they are entirely univariate in nature – that is, they model the conditional variance of each series entirely independently of all other series. This is potentially an important limitation for two reasons. First, to the extent that there may be 'volatility spillovers' between markets or assets (a tendency for volatility to change in one market or asset following a change in the volatility of another), the univariate model will be misspecified. For instance, using a multivariate model will allow us to determine whether the volatility in one market leads or lags the volatility in others.

Second, it is often the case in finance that the covariances between series are of interest, as well as the variances of the individual series themselves. The calculation of hedge ratios, portfolio value at risk estimates, CAPM betas, and so on, all require covariances as inputs.

Multivariate GARCH models can potentially overcome both of these deficiencies with their univariate counterparts. Multivariate extensions to GARCH models can be used to forecast the volatilities of the component series, just as with univariate models and since the volatilities of financial time series often move together, a joint approach to modelling may be more efficient than treating each separately. In addition, because multivariate models give estimates for the conditional covariances as well as the conditional variances, they have a number of other potentially useful applications.

Several papers have investigated the forecasting ability of various models incor-porating correlations. Siegel (1997), for example, finds that implied correlation forecasts from traded options encompass all information embodied in the historical returns (although he does not consider EWMA- or GARCH-based models). Walter and Lopez (2000), on the other hand, find that implied correlation is generally less useful for predicting the future correlation between the underlying assets' returns than forecasts derived from GARCH models. Finally, Gibson and Boyer (1998) find that a diagonal GARCH and a Markov switching approach provide better correlation forecasts than simpler models in the sense that the latter produce smaller profits when the forecasts are employed in a trading strategy.

9.21 Covariance Modelling and Forecasting in Finance: Some Examples

9.21.1 The Estimation of Conditional Betas

The CAPM beta for asset *i* is defined as the ratio of the covariance between the market portfolio return and the asset return, to the variance of the market portfolio return. Betas are typically constructed using a set of historical data on market variances and covariances. However, like most other problems in finance, beta estimation conducted in this fashion is backward-looking, when investors should really be concerned with the beta that will prevail in the future over the time that the investor is considering holding the asset. Multivariate GARCH models provide a simple method for estimating conditional (or time-varying) betas. Then forecasts of the covariance between the asset and the market portfolio returns and forecasts of the variance of the market portfolio are made from the model, so that the beta is a forecast, whose value will vary over time

$$\beta_{i,t} = \frac{\sigma_{im,t}}{\sigma_{m,t}^2} \quad \textbf{(9.90)}$$

where $\beta_{i,t}$ is the time-varying beta estimate at time t for stock i, $\sigma_{im,t}$ is the covariance between market returns and returns to stock i at time t and $\sigma_{m,t}^2$ is the variance of the market return at time t.

9.21.2 Dynamic Hedge Ratios

Although there are many techniques available for reducing and managing risk, the simplest and perhaps the most widely used, is hedging with futures contracts. A hedge is achieved by taking opposite positions in spot and futures markets simultaneously, so that any loss sustained from an adverse price movement in one market should to some degree be offset by a favourable price movement in the other. The ratio of the number of units of the futures asset that are purchased relative to the number of units of the spot asset is known as the *hedge ratio*. Since risk in this context is usually measured as the volatility of portfolio returns, an intuitively plausible strategy might be to choose that hedge ratio which minimises the variance of the returns of a portfolio containing the spot and futures position; this is known as the *optimal hedge ratio*. The optimal value of the hedge ratio may be determined in the usual way, following Hull (2017) by first defining:

ΔS = change in spot price S, during the life of the hedge ΔF = change in futures price, F, during the life of the hedge σ_s = standard deviation of $\Delta S \sigma_F$ = standard deviation of $\Delta F p$ = correlation coefficient between ΔS and $\Delta F h$ = hedge ratio

For a short hedge (i.e., long in the asset and short in the futures contract), the change in the value of the hedger's position during the life of the hedge will be given by ($\Delta S - h\ \Delta F$), while for a long hedge, the appropriate expression will be ($h\Delta F - \Delta S$).

The variances of the two hedged portfolios (long spot and short futures or long futures and short spot) are the same. These can be obtained from

$$\mathrm{var}(h\Delta F - \Delta S)$$

Remembering the rules for manipulating the variance operator, this can be written

$$\mathrm{var}(\Delta S) + \mathrm{var}(h\Delta F) - 2\mathrm{cov}(\Delta S, h\Delta F)$$

or

$$\mathrm{var}(\Delta S) + h^2\mathrm{var}(\Delta F) - 2h\mathrm{cov}(\Delta S, \Delta F)$$

Hence the variance of the change in the value of the hedged position is given by

$$v = \sigma_s^2 + h^2\sigma_F^2 - 2hp\sigma_s\sigma_F \qquad \textbf{(9.91)}$$

Minimising this expression w.r.t. *h* would give

$$h = p\frac{\sigma_s}{\sigma_F} \qquad \textbf{(9.92)}$$

Again, according to this formula, the optimal hedge ratio is time-invariant, and would be calculated using historical data. However, what if the standard deviations are changing over time? The standard deviations and the correlation between movements in the spot and futures series could be forecast from a multivariate GARCH model, so that the expression above is replaced by

$$h_t = p_t\frac{\sigma_{s,t}}{\sigma_{F,t}} \qquad \textbf{(9.93)}$$

Various models are available for covariance or correlation forecasting, and several will be discussed below, which are grouped into simple models, multivariate GARCH models, and specific correlation models.

9.22 Simple Covariance Models

9.22.1 Historical Covariance and Correlation

In exactly the same fashion as for volatility, the historical covariance or correlation between two series can be calculated in the standard way using a set of historical data.

9.22.2 Implied Covariance Models

Implied covariances can be calculated using options whose payoffs are dependent on more than one underlying asset. The relatively small number of such options that exist limits the circumstances in which implied covariances can be calculated. Examples include rainbow options, ‘crack-spread’ options for different grades of oil, and currency options. In the latter case, the implied variance of the cross-currency returns *xy* is given by

$$\tilde{\sigma}^2(xy) = \tilde{\sigma}^2(x) + \tilde{\sigma}^2(y) - 2\tilde{\sigma}(x, y) \quad \textbf{(9.94)}$$

where $\tilde{\sigma}^2(x)$ and $\tilde{\sigma}^2(y)$ are the implied variances of the *x* and *y* returns, respectively, and $\tilde{\sigma}(x, y)$ is the implied covariance between *x* and *y*. By substituting the observed option implied volatilities of the three currencies into equation (9.94), the implied covariance is obtained via

$$\tilde{\sigma}(x, y) = \frac{\tilde{\sigma}^2(x) + \tilde{\sigma}^2(y) - \tilde{\sigma}^2(xy)}{2} \quad \textbf{(9.95)}$$

So, for instance, if the implied covariance between USD/DEM and USD/JPY is of interest, then the implied variances of the returns of USD/DEM and USD/JPY, as well as the returns of the cross-currency DEM/JPY, are required so as to obtain the implied covariance using equation (9.94).

9.22.3 Exponentially Weighted Moving Average Model for Covariances

Again, as for the case of single series volatility modelling, a EWMA specification is available that gives more weight in the calculation of covariance to recent observations than the estimate based on the simple average. The EWMA model estimates for variances and covariances at time *t* in the bivariate setup with two returns series *x* and *y* may be written as

$$h_{ij,t} = \lambda h_{ij,t-1} + (1 - \lambda)x_{t-1}y_{t-1} \quad \textbf{(9.96)}$$

where $i \neq j$ for the covariances and $i = j$; $x = y$ for the variance specifications. As for the univariate case, the fitted values for *h* also become the forecasts for subsequent periods. $\lambda (0 < \lambda < 1)$ again denotes the decay factor determining the relative weights attached to recent versus less recent observations. this parameter could be estimated (for example, by maximum likelihood), but is often set arbitrarily (– for example, Riskmetrics use a decay factor of 0.97 for monthly data but 0.94 when the data are of daily frequency).

This equation can be rewritten as an infinite order function of only the returns by successively substituting out the covariances

$$h_{ij,t} = (1-\lambda)\sum_{i=0}^{\infty}\lambda^{i}x_{t-i}y_{t-i} \quad \textbf{(9.97)}$$

While the EWMA model is probably the simplest way to allow for time-varying variances and covariances, the model is a restricted version of an integrated GARCH (IGARCH) specification, and it does not guarantee the fitted variance-covariance matrix to be positive definite. As a result of the parallel with IGARCH, EWMA models also cannot allow for the observed mean reversion in the volatilities or covariances of asset returns that is particularly prevalent at lower frequencies of observation.

9.23 Multivariate GARCH Models

Multivariate GARCH models are in spirit very similar to their univariate counterparts, except that the former also specify equations for how the covariances move over time and are therefore by their nature inherently more complex to specify and estimate. Several different multivariate GARCH formulations have been proposed in the literature, the most popular of which are the *VECH*, the diagonal *VECH* and the *BEKK* models. Each of these and several others is discussed in turn below; for a more detailed discussion, see Kroner and Ng (1998). In each case, there are N assets, whose return variances and covariances are to be modelled.

9.23.1 The VECH model

As with univariate GARCH models, the conditional mean equation may be parameterised in any way desired, although it is worth noting that, since the conditional variances are measured about the mean, misspecification of the latter is likely to imply misspecification of the former. To introduce some notation, suppose, that y_t $(y_{1t}\ y_{2t}\ \dots\ y_{Nt})$, is an $N \times 1$ vector of time-series observations, C is an $N(N+1)/2$ column vector of conditional variance and covariance intercepts, and A and B are square parameter matrices of order $N(N+1)/2$. A common specification of the *VECH* model, initially due to Bollerslev, Engle and Wooldridge (1988), is

$$VECH(H_t) = C + AVECH(\Xi_{t-1}\Xi'_{t-1}) + BVECH(H_{t-1})$$
$$\Xi_t|\psi_{t-1} \sim N(0, H_t), \qquad \textbf{(9.98)}$$

where H_t is a $N \times N$ conditional variance–covariance matrix, Ξ_t is a $N \times 1$ innovation (disturbance) vector, ψ_{t-1} represents the information set at time $t - 1$, and *VECH* (·) denotes the column-stacking operator applied to the upper portion of the symmetric matrix. In the bivariate case (i.e., $N = 2$), C will be a 3×1 parameter vector, and A and B will be 3×3 parameter matrices.

The unconditional variance matrix for the *VECH* will be given by $C[I - A - B]^{-1}$, where I is an identity matrix of order $N(N + 1)/2$. Stationarity of the *VECH* model requires that the eigenvalues of $[A + B]$ are all less than one in absolute value.

In order to gain a better understanding of how the *VECH* model works, the elements for $N = 2$ are written out below. Define

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{bmatrix}, \Xi_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, C = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix},$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix},$$

The *VECH* operator takes the 'upper triangular' portion of a matrix, and stacks each element into a vector with a single column. For example, in the case of $VECH(H_t)$, this becomes

$$VECH(H_t) = \begin{bmatrix} h_{11t} \\ h_{22t} \\ h_{12t} \end{bmatrix}$$

where h_{iit} represent the conditional variances at time t of the two-asset return series ($i = 1, 2$) used in the model, and h_{ijt} ($i \neq j$) represent the conditional covariances between the asset returns. In the case of $VECH(\Xi_t\Xi'_t)$, this can be expressed as

$$
\begin{aligned}
VECH(\Xi_t \Xi_t') &= VECH\left(\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} [u_{1t} \quad u_{2t}]\right) \\
&= VECH\begin{pmatrix} u_{1t}^2 & u_{1t}u_{2t} \\ u_{1t}u_{2t} & u_{2t}^2 \end{pmatrix} \\
&= \begin{bmatrix} u_{1t}^2 \\ u_{2t}^2 \\ u_{1t}u_{2t} \end{bmatrix}
\end{aligned}
$$

The *VECH* model in full is given by

$$
\begin{aligned}
h_{11t} = c_{11} + a_{11}u_{1t-1}^2 + a_{12}u_{2t-1}^2 + a_{13}u_{1t-1}u_{2t-1} + b_{11}h_{11t-1} \\
+ b_{12}h_{22t-1} + b_{13}h_{12t-1}
\end{aligned} \tag{9.99}
$$

$$
\begin{aligned}
h_{22t} = c_{21} + a_{21}u_{1t-1}^2 + a_{22}u_{2t-1}^2 + a_{23}u_{1t-1}u_{2t-1} + b_{21}h_{11t-1} \\
+ b_{22}h_{22t-1} + b_{23}h_{12t-1}
\end{aligned} \tag{9.100}
$$

$$
\begin{aligned}
h_{12t} = c_{31} + a_{31}u_{1t-1}^2 + a_{32}u_{2t-1}^2 + a_{33}u_{1t-1}u_{2t-1} + b_{31}h_{11t-1} \\
+ b_{32}h_{22t-1} + b_{33}h_{12t-1}
\end{aligned} \tag{9.101}
$$

Thus, it is clear that the conditional variances and conditional covariances depend on the lagged values of all of the conditional variances of, and conditional covariances between, all of the asset returns in the series, as well as the lagged squared errors and the error cross-products. This unrestricted model is highly parameterised, and it is challenging to estimate. For N = 2 there are 21 parameters (C has 3 elements, A and B each have 9 elements), while for N = 3 there are 78, and N = 4 implies 210 parameters!

9.23.2 The Diagonal VECH Model

As the number of assets employed increases, estimation of the *VECH* model can quickly become infeasible. Hence the *VECH* model's conditional variance–covariance matrix has been restricted to the form developed by Bollerslev, Engle and Wooldridge (1988), in which A and B are assumed to be diagonal. This restriction implies that there are no direct volatility spillovers from one series to another, which considerably reduces the number of parameters to be estimated to nine in the bivariate case (now A and B each have three elements) and 18 for a trivariate system (i.e., if N = 3). The model, known as a diagonal *VECH*, is now characterised by

$$
h_{ij,t} = \omega_{ij} + \alpha_{ij}u_{i,t-1}u_{j,t-1} + \beta_{ij}h_{ij,t-1} \quad \text{for } i,j = 1,2
$$

(9.102)

where ω_{ij}, α_{ij} and β_{ij} are parameters.

The diagonal *VECH* multivariate GARCH model could also be expressed as an infinite order multivariate ARCH model, where the covariance is expressed as a geometrically declining weighted average of past cross products of unexpected returns, with recent observations carrying higher weights. An alternative solution to the dimensionality problem would be to use orthogonal GARCH (see, for example, Van der Weide, 2002) or factor GARCH models (see Engle, Ng and Rothschild, 1990). A disadvantage of the *VECH* model is that there is no guarantee of a positive semi-definite covariance matrix.

A variance–covariance or correlation matrix must always be 'positive semi-definite', and in the case where all the returns in a particular series are all the same so that their variance is zero is disregarded, then the matrix will be positive definite. Among other things, this means that the variance–covariance matrix will have all positive numbers on the leading diagonal, and will be symmetrical about this leading diagonal. These properties are intuitively appealing as well as important from a mathematical point of view, for variances can never be negative, and the covariance between two series is the same irrespective of which of the two series is taken first, and positive definiteness ensures that this is the case.

A positive definite correlations matrix is also important for many applications in finance – for example, from a risk management point of view. It is this property which ensures that, whatever the weight of each series in the asset portfolio, an estimated value-at-risk is always positive. Fortunately, this desirable property is automatically a feature of time-invariant correlations matrices which are computed directly using actual data. An anomaly arises when either the correlation matrix is estimated using a non-linear optimisation procedure (as multivariate GARCH models are), or when modified values for some of the correlations are used by the risk manager. The resulting modified correlation matrix may or may not be positive definite, depending on the values of the correlations that are put in, and the values of the remaining correlations. If, by chance, the matrix is not positive definite, the upshot is that for some weightings of the individual assets in the portfolio, the estimated portfolio variance could be negative.

9.23.3 The BEKK model

The *BEKK* model (Engle and Kroner, 1995) addresses the difficulty with *VECH* of ensuring that the H matrix is always positive definite.[1] It is represented by

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi'_{t-1}B \quad \textbf{(9.103)}$$

where A, and B are $N \times N$ matrices of parameters and W is an upper triangular matrix of parameters. The positive definiteness of the covariance matrix is ensured owing to the quadratic nature of the terms on the equation's RHS.

9.23.4 Model Estimation for Multivariate GARCH

Under the assumption of conditional normality, the parameters of the multivariate GARCH models of any of the above specifications can be estimated by maximising the log-likelihood function

$$\ell(\theta) = -\frac{TN}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\ln|H_t| + \Xi_t' H_t^{-1}\Xi_t\right) \quad \textbf{(9.104)}$$

where θ denotes all the unknown parameters to be estimated, N is the number of assets (i.e., the number of series in the system) and T is the number of observations and all other notation is as above. The maximum-likelihood estimate for θ is asymptotically normal, and thus traditional procedures for statistical inference are applicable. Further details on maximum-likelihood estimation in the context of multivariate GARCH models are beyond the scope of this book. But suffice to say that the additional complexity and extra parameters involved compared with univariate models make estimation a computationally more difficult task, although the principles are essentially the same.

9.24 Direct Correlation Models

The *VECH* and *BEKK* models specify the dynamics of the covariances between a set of series, and the correlations between any given pair of series at each point in time can be constructed by dividing the conditional covariances by the product of the conditional standard deviations. A subtly different approach would be to model the dynamics for the correlations directly – Bauwens, Laurent and Rombouts (2006) term these 'non-linear combinations of univariate GARCH models' for reasons that will become apparent in the following sub-section.

9.24.1 The Constant Correlation Model

An alternative method for reducing the number of parameters in the MGARCH framework is to require the correlations between the disturbances, ϵ_t, (or equivalently between the observed variables, y_t) to be fixed through time. Thus, although the conditional covariances are not fixed, they are tied to the variances as proposed in the constant conditional correlation (CCC) model due to Bollerslev (1990). The conditional variances in the fixed correlation model are identical to those of a set of univariate GARCH specifications (although they are estimated jointly)

$$h_{ii,t} = c_i + a_i \epsilon_{i,t-i}^2 + b_i h_{ii,t-1}, \qquad i = 1, \ldots, N \tag{9.105}$$

The off-diagonal elements of H_t, $h_{ij,t}$ ($i \neq j$), are defined indirectly via the correlations, denoted ρ_{ij}

$$h_{ij,t} = \rho_{ij} h_{ii,t}^{1/2} h_{jj,t}^{1/2}, \qquad i,j = 1, \ldots, N, i < j \tag{9.106}$$

Is it empirically plausible to assume that the correlations are constant through time? Several tests of this assumption have been developed, including a test based on the information matrix due to Bera and Kim (2002) and a Lagrange Multiplier test due to Tse (2000). The conclusions reached appear dependent on which test is used, but there seems to be non-negligible evidence against constant correlations, particularly in the context of stock returns.

9.24.2 The Dynamic Conditional Correlation Model

Several different formulations of the dynamic conditional correlation (DCC) model are available, but a popular specification is due to Engle (2002). The model is related to the CCC formulation described above, but where the correlations are allowed to vary over time. Define the variance–covariance matrix, H_t, as

$$H_t = D_t R_t D_t \tag{9.107}$$

where D_t is a diagonal matrix containing the conditional standard deviations (i.e., the square roots of the conditional variances from univariate GARCH model estimations on each of the N individual series) on the leading diagonal; R_t is the conditional correlation matrix. Forcing R_t to be time-invariant would lead back to the constant conditional correlation model.

Numerous explicit parameterisations of R_t are possible, including an exponential smoothing approach discussed in Engle (2002). More generally, a model of the MGARCH form could be specified as

$$Q_t = S \circ (\iota\iota' - A - B) + A \circ u_{t-1}u'_{t-1} + B \circ Q_{t-1} \quad \textbf{(9.108)}$$

where S is the unconditional correlation matrix of the vector of standardised residuals (from the first stage estimation – see below), $u_t = D_t^{-1}\epsilon_t$. ι is a vector of ones, and Q_t is an $N \times N$ symmetric positive definite variance-covariance matrix. $\circ$ denotes the *Hadamard* or element-by-element matrix multiplication procedure. This specification for the intercept term simplifies estimation and reduces the number of parameters to be estimated, but is not necessary. Engle (2002) proposes a GARCH-esque formulation for dynamically modelling Q_t with the conditional correlation matrix, R_t, then constructed as

$$R_t = diag\{Q_t^*\}^{-1} Q_t diag\{Q_t^*\}^{-1} \quad \textbf{(9.109)}$$

where $diag(\cdot)$ denotes a matrix comprising the main diagonal elements of $(\cdot)$ and Q^* is a matrix that takes the square roots of each element in Q. This operation is effectively taking the covariances in Q_t and dividing them by the product of the appropriate standard deviations in Q_t^* to create a matrix of correlations.

A slightly different form of the DCC was proposed by Tse and Tsui (2002), and equation (9.108) could also be simplified by specifying A and B each as single scalars so that all the conditional correlations would follow the same process.

The model may be estimated in one single stage using maximum likelihood, although this will still be a difficult exercise in the context of large systems. Consequently, Engle advocates a two-stage estimation procedure where each variable in the system is first modelled separately as a univariate GARCH process. A joint log-likelihood function for this stage could be constructed, which would simply be the sum (over N) of all of the log-likelihoods for the individual GARCH models. Then, in the second stage, the conditional likelihood is maximised with respect to any unknown parameters in the correlation matrix. The log-likelihood function for the second stage estimation will be of the form

$$\ell(\theta_2|\theta_1) = \sum_{t=1}^{T} \left(\ln|R_t| + u_t' R_t^{-1} u_t \right) \quad \textbf{(9.110)}$$

where θ_1 denotes all the unknown parameters that were estimated in the first stage and θ_2 denotes all those to be estimated in the second stage. Estimation using this two-step procedure will be consistent but inefficient as a result of any parameter uncertainty from the first stage being carried through to the second.

9.25 Extensions to the Basic Multivariate GARCH Model

Numerous extensions to the univariate specification have been proposed, and many of these carry over to the multivariate case. For example, conditional variance or covariance terms can be included in the conditional mean equation (see Bollerslev, Engle and Wooldridge, 1988, for instance). In the context of financial applications, where the y_t are returns, the parameters on these variables can be loosely interpreted as risk premia.

9.25.1 Asymmetric Multivariate GARCH

Asymmetric models have become very popular in empirical applications, where the conditional variances and/or covariances are permitted to react differently to positive and negative innovations of the same magnitude. In the multivariate context, this is usually achieved in the Glosten, Jagannathan and Runkle (1993) framework, rather than the EGARCH specification of Nelson (1991). Kroner and Ng (1998), for example, suggest the following extension to the BEKK formulation (with obvious related modifications for the *VECH* or diagonal *VECH* models)

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi_{t-1}'B + D'z_{t-1}z_{t-1}'D \quad \textbf{(9.111)}$$

where z_{t-1} is an N-dimensional column vector with elements taking the value $-\epsilon_{t-1}$ if the corresponding element of ϵ_{t-1} is negative and zero otherwise. The asymmetric properties of time-varying covariance matrix models are analysed by Kroner and Ng (1998), who identify three possible forms of asymmetric behaviour. First, the covariance matrix displays own variance asymmetry if the conditional variance of one series is affected by the sign of the innovation in that series. Second, the covariance matrix displays cross variance asymmetry if the conditional variance of one series is affected by the sign of the innovation of another series. Finally, if the conditional covariance is sensitive to the sign of the innovation in return for either series, then the model is said to display covariance asymmetry.

9.25.2 Alternative Distributional Assumptions

As was the case for stochastic volatility and univariate GARCH models, an assumption of (multivariate) conditional normality cannot generate sufficiently fat tails to accurately model the distributional properties of financial data. A better approximation to the actual distributions of (especially financial) time series can be obtained using a Student's *t* distribution. Such a model can still be estimated using maximum likelihood but with a different (and more complex) likelihood function. The standard formulation will involve estimating, as part of the process, a single degree of freedom parameter which applies to all of the series in the system. An additional potential drawback of this approach is that the tail fatness embodied in the degrees of freedom parameter is fixed over time. Brooks *et al.* (2005) propose a model where both of these limitations are removed. However, some identifying restrictions are still required. A further issue is the extent to which the unconditional distribution of the shocks is skewed. If this is the case, then a model based on the Student's *t* will be inadequate, and an alternative such as the multivariate skew Student's *t* of Bauwens and Laurent (2002) must be used.

Although many other extensions of the basic models may be conceived of, such as periodic or seasonal MGARCH, the range of specifications employed in the existing literature is narrower than for the corresponding univariate models. A major drawback for even the more parsimonious of the models above is that they are too highly parameterised, and yet many potential applications in economics and finance are in the context of high dimensional systems (such as asset allocation among a number of stocks). Thus, an important innovation was the development of orthogonal and factor models referenced above. Both have the same fundamental idea that by forcing some structure on the variance-covariance matrix, a simplification can be achieved.

9.26 A Multivariate GARCH Model for the CAPM

with Time-Varying Covariances

Bollerslev, Engle and Wooldridge (1988) estimate a multivariate GARCH model for returns to US Treasury bills, gilts and stocks. The data employed comprised calculated quarterly excess holding period returns for six-month US Treasury bills, twenty-year US Treasury bonds and a Center for Research in Security Prices record of the return on the New York Stock Exchange (NYSE) value-weighted index. The data run from 1959Q1 to 1984Q2 – a total of 102 observations.

A multivariate GARCH-M model of the diagonal *VECH* type is employed, with coefficients estimated by maximum likelihood, and the Berndt *et al.* (1974) algorithm is used. The coefficient estimates are easiest presented in the following equations for the conditional mean and variance equations, respectively

$$\begin{vmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{vmatrix} = \begin{vmatrix} \underset{(0.032)}{0.070} \\ \underset{(1.030)}{-4.342} \\ \underset{(0.710)}{-3.117} \end{vmatrix} + \underset{(0.160)}{0.499} \sum_j \omega_{jt-1} \begin{vmatrix} h_{1jt} \\ h_{2jt} \\ h_{3jt} \end{vmatrix} + \begin{vmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{vmatrix} \tag{9.112}$$

$$\begin{vmatrix} h_{11t} \\ h_{12t} \\ h_{22t} \\ h_{13t} \\ h_{23t} \\ h_{33t} \end{vmatrix} = \begin{vmatrix} \underset{(0.004)}{0.011} \\ \underset{(0.062)}{0.176} \\ \underset{(6.372)}{13.305} \\ \underset{(0.009)}{0.018} \\ \underset{(2.820)}{5.143} \\ \underset{(1.466)}{2.083} \end{vmatrix} + \begin{vmatrix} \underset{(0.105)}{0.445\varepsilon_{1t-1}^2} \\ \underset{(0.092)}{0.233\varepsilon_{1t-1}\varepsilon_{2t-1}} \\ \underset{(0.113)}{0.188\varepsilon_{2t-1}^2} \\ \underset{(0.132)}{0.197\varepsilon_{1t-1}\varepsilon_{3t-1}} \\ \underset{(0.093)}{0.165\varepsilon_{2t-1}\varepsilon_{3t-1}} \\ \underset{(0.066)}{0.078\varepsilon_{3t-1}^2} \end{vmatrix} + \begin{vmatrix} \underset{(0.056)}{0.466h_{11t-1}} \\ \underset{(0.052)}{0.598h_{12t-1}} \\ \underset{(0.215)}{0.441h_{22t-1}} \\ \underset{(0.361)}{-0.362h_{13t-1}} \\ \underset{(0.338)}{-0.348h_{23t-1}} \\ \underset{(0.333)}{0.469h_{33t-1}} \end{vmatrix} \tag{9.113}$$

Source: Bollerslev, Engle and Wooldridge (1988). Reprinted with the permission of University of Chicago Press.

where y_{jt} are the returns, ω_{jt-1} are a set vector of value weights at time $t - 1$, $i = 1, 2, 3$, refers to bills, bonds and stocks, respectively and standard errors are given in parentheses. Consider now the implications of the signs, sizes and significances of the coefficient estimates in equations (9.112) and (9.113). The coefficient of 0.499 in the conditional mean equation gives an aggregate measure of relative risk aversion, also interpreted as representing the market trade-off between return and risk. This conditional variance-in-mean coefficient gives the required additional return as compensation for taking an additional unit of variance (risk). The intercept coefficients in the conditional mean equation for bonds and stocks are very negative and highly statistically significant. The authors argue that this is to be expected since favourable tax treatments for investing in longer-term assets encourages investors to hold them even at relatively low rates of return.

The dynamic structure in the conditional variance and covariance equations is strongest for bills and bonds, and very weak for stocks, as indicated by their respective statistical significances. In fact, none of the parameters in the conditional variance or covariance equations for the stock return equations is significant at the 5% level. The unconditional covariance between bills and bonds is positive, while that between bills and stocks, and between bonds and stocks, is negative. This arises since, in the latter two cases, the lagged conditional covariance parameters are negative and larger in absolute value than those of the corresponding lagged error cross-products.

Finally, the degree of persistence in the conditional variance (given by $\alpha_1 + \beta$), which embodies the degree of clustering in volatility, is relatively large for the bills equation, but surprisingly small for bonds and stocks, given the results of other relevant papers in this literature.

9.27 Estimating a Time-Varying Hedge Ratio for FTSE Stock Index Returns

A paper by Brooks, Henry and Persand (2002) compared the effectiveness of hedging on the basis of hedge ratios derived from various multivariate GARCH specifications and other, simpler techniques. Some of their main results are discussed below.

9.27.1 Background

There has been much empirical research into the calculation of optimal hedge ratios. The general consensus is that the use of multivariate GARCH (MGARCH) models yields superior performances, evidenced by lower portfolio volatilities, than either time-invariant or rolling OLS hedges. Cecchetti, Cumby and Figlewski (1988), Myers and Thompson (1989) and Baillie and Myers (1991), for example, argue that commodity prices are characterised by time-varying covariance matrices. As news about spot and futures prices arrives to the market in discrete bunches, the conditional covariance matrix, and hence the optimal hedging ratio, becomes time-varying. Baillie and Myers (1991) and Kroner and Sultan (1993), *inter alia*, employ MGARCH models to capture time-variation in the covariance matrix and to estimate the resulting hedge ratio.

9.27.2 Notation

Let S_t and F_t represent the logarithms of the stock index and stock index futures prices, respectively. The actual return on a spot position held from time $t-1$ to t is $\Delta S_t = S_t - S_{t-1}$ similarly, the actual return on a futures position is $\Delta F_t = F_t - F_{t-1}$. However at time $t-1$ the expected return, $E_{t-1}(R_t)$, of the portfolio comprising one unit of the stock index and β units of the futures contract may be written as

$$E_{t-1}(R_t) = E_{t-1}(\Delta S_t) - \beta_{t-1}E_{t-1}(\Delta F_t) \tag{9.114}$$

where β_{t-1} is the hedge ratio determined at time $t-1$, for employment in period t. The variance of the expected return, $h_{p,t}$, of the portfolio may be written as

$$h_{p,t} = h_{s,t} + \beta_{t-1}^2 h_{F,t} - 2\beta_{t-1}h_{SF,t} \tag{9.115}$$

where $h_{p,t}$, $h_{s,t}$ and $h_{F,t}$ represent the conditional variances of the portfolio and the spot and futures positions, respectively and $h_{SF,t}$ represents the conditional covariance between the spot and futures position. β_{t-1}^*, the optimal number of futures contracts in the investor's portfolio, i.e., the optimal hedge ratio, is given by

$$\beta_{t-1}^* = -\frac{h_{SF,t}}{h_{F,t}} \tag{9.116}$$

If the conditional variance–covariance matrix is time-invariant (and if S_t and F_t are not cointegrated) then an estimate of β^*, the constant optimal hedge ratio, may be obtained from the estimated slope coefficient b in the regression

$$\Delta S_t = a + b\Delta F_t + u_t \tag{9.117}$$

The OLS estimate of the optimal hedge ratio could be given by $b = h_{SF}/h_F$.

9.27.3 Data and Results

The data employed in the Brooks, Henry and Persand (2002) study comprise 3,580 daily observations on the FTSE 100 stock index and stock index futures contract spanning the period 1 January 1985–9 April 1999. Several approaches to estimating the optimal hedge ratio are investigated.

The hedging effectiveness is first evaluated in-sample, that is, where the hedges are constructed and evaluated using the same set of data. The out-of-sample hedging effectiveness for a one-day hedging horizon is also investigated by forming one-step-ahead forecasts of the conditional variance of the futures series and the conditional covariance between the spot and futures series. These forecasts are then translated into hedge ratios using equation (9.116). The hedging performance of a *BEKK* formulation is examined, and also a *BEKK* model including asymmetry terms (in the same style as GJR models). The returns and variances for the various hedging strategies are presented in Table 9.5.

Table 9.5 Hedging effectiveness: summary statistics for portfolio returns

In-sample				
(1)	Unhedged $\beta = 0$ (2)	Naive hedge $\beta = -1$ (3)	Symmetric time-varying hedge $\beta_t = \frac{h_{FS,t}}{h_{F,t}}$ (4)	Asymmetric time-varying hedge $\beta_t = \frac{h_{FS,t}}{h_{F,t}}$ (5)
Return	0.0389	−0.0003	0.0061	0.0060
	{2.3713}	{−0.0351}	{0.9562}	{0.9580}
Variance	0.8286	0.1718	0.1240	0.1211
Out-of-sample				
	Unhedged $\beta = 0$	Naive hedge $\beta = -1$	Symmetric time-varying hedge $\beta_t = \frac{h_{FS,t}}{h_{F,t}}$	Asymmetric time-varying hedge $\beta_t = \frac{h_{FS,t}}{h_{F,t}}$
Return	0.0819	−0.0004	0.0120	0.0140
	{1.4958}	{0.0216}	{0.7761}	{0.9083}
Variance	1.4972	0.1696	0.1186	0.1188

Note: t-ratios displayed as {.}.

Source: Brooks, Henry and Persand (2002).

The simplest approach, presented in column (2), is that of no hedge at all. In this case, the portfolio simply comprises a long position in the cash market. Such an approach is able to achieve significant positive returns in sample, but with a large variability of portfolio returns. Although none of the alternative strategies generate returns that are significantly different from zero, either in-sample or out-of-sample, it is clear from columns (3)–(5) of Table 9.5 that any hedge generates significantly less return variability than none at all.

The ‘naive’ hedge, which takes one short futures contract for every spot unit, but does not allow the hedge to time-vary, generates a reduction in variance of the order of 80% in-sample and nearly 90% out-of-sample relative to the unhedged position. Allowing the hedge ratio to be time-varying and determined from a symmetric multivariate GARCH model leads to a further reduction as a proportion of the unhedged variance of 5% and 2% for the in-sample and holdout sample, respectively. Allowing for an asymmetric response of the conditional variance to positive and negative shocks yields a very modest reduction in variance (a further 0.5% of the initial value) in-sample, and virtually no change out-of-sample.

Figure 9.5 graphs the time-varying hedge ratio from the symmetric and asymmetric MGARCH models (source: Brooks, Henry and Persand, 2002). The optimal hedge ratio is never greater than 0.96 futures contracts per index contract, with an average value of 0.82 futures contracts sold per long index contract. The variance of the estimated optimal hedge ratio is 0.0019. Moreover the optimal hedge ratio series obtained through the estimation of the asymmetric GARCH model appears stationary. An ADF test of the null hypothesis $\beta^*_{t-1} \sim \mathrm{I}(1)$ (i.e., that the optimal hedge ratio from the asymmetric *BEKK* model contains a unit root) was strongly rejected by the data (ADF statistic = −5.7215, 5% Critical value = −2.8630). The time-varying hedge requires the sale (purchase) of fewer futures contracts per long (short) index contract and hence would save the firm wishing to hedge a short exposure money relative to the time-invariant hedge. One possible interpretation of the better performance of the dynamic strategies over the naive hedge is that the dynamic hedge uses short-run information, while the naive hedge is driven by long-run considerations and an assumption that the relationship between spot and futures price movements is 1:1.

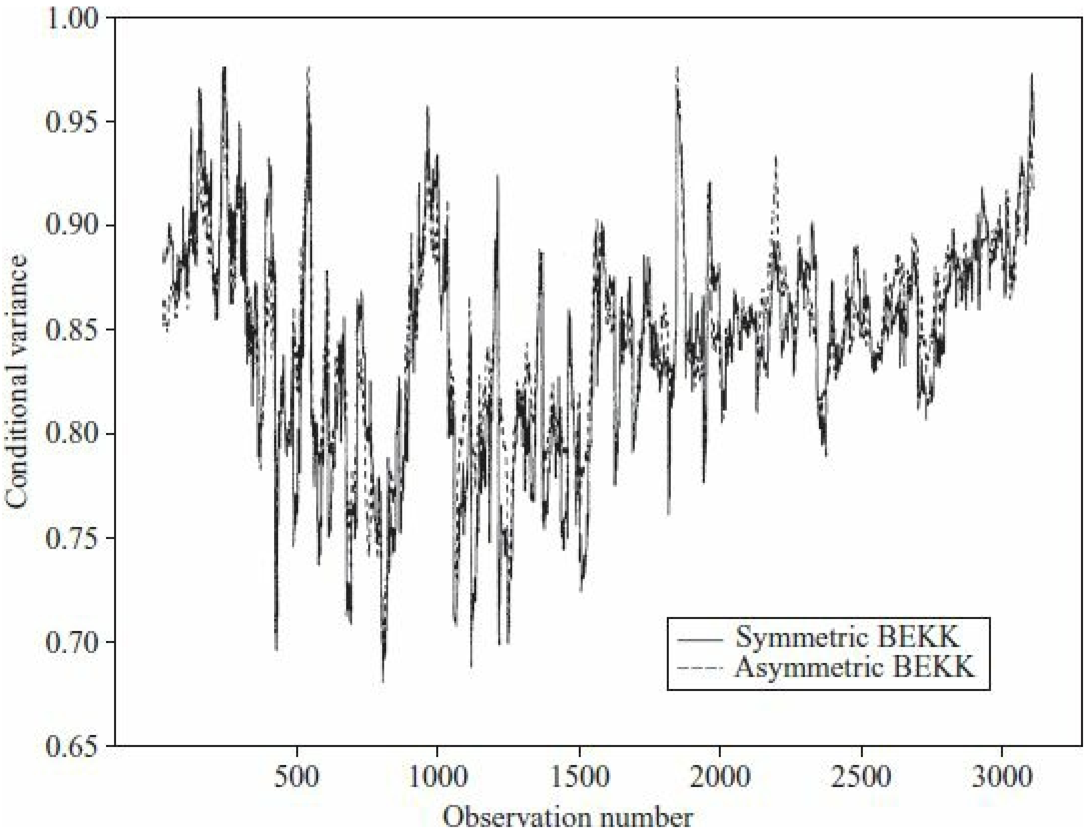

Figure 9.5 Time-varying hedge ratios derived from symmetric and asymmetric *BEKK* models for FTSE returns

Source. Brooks, Henry and Persand (2002).

Brooks, Henry and Persand also investigate the hedging performances of the various models using a modern risk management approach. They find, once again, that the time-varying hedge results in a considerable improvement, but that allowing for asymmetries results in only a very modest incremental reduction in hedged portfolio risk.

9.28 Multivariate Stochastic Volatility Models

As in the univariate case, while the term ‘stochastic volatility’ is commonly used to describe models from the multivariate GARCH family, strictly they do not fit well under this umbrella because the conditional variance and covariance equations are deterministic given the information set up to the previous period. That is, there is no additional source of noise in the conditional variance (or covariance) equation of a multivariate GARCH model.

The multivariate stochastic volatility (MSV) model was initially proposed by Harvey, Ruiz and Shephard (1994) and the notation here will closely follow theirs. Let y_t be the elements of an $N \times 1$ vector of observations at time t on a series i, with time-varying variance σ_i^2, defined as

$$y_{it} = \epsilon_{it}(\exp\{h_{it}\})^{1/2}, \qquad i = 1, \ldots, N; t = 1, \ldots, T \quad \textbf{(9.118)}$$

where $\epsilon = (\epsilon_{1t}, \ldots, \epsilon_{Nt})$ is a vector of disturbances with zero mean and covariance matrix Σ_ϵ and where

$$h_{it} = \ln(\sigma_{it}^2) \quad \textbf{(9.119)}$$

This covariance matrix, Σ_ϵ is defined to have unity on the leading diagonal (and it is therefore also a correlation matrix), while its off-diagonal elements are denoted ρ_{ij}.

Under the stochastic volatility model, the h_{it} can be specified to evolve as an autoregressive (AR) process of order P

$$h_{it} = \gamma_i + \sum_{p=1}^{P} \psi_{ip} h_{i,t-p} + \eta_{it} \qquad i = 1, \ldots, N \quad \textbf{(9.120)}$$

$\eta_t = (\eta_{1t}, \ldots, \eta_{Nt})$ is a vector of disturbances to the conditional variance having zero mean and covariance matrix Σ_η. It is usually further assumed that ϵ_{it} and η_{it} are mutually independent and that each is multivariate normally distributed. Usually, $P = 1$ is deemed sufficient so that the variance dynamics for each series in the system are AR(1). Moving average terms or even exogenous variables could be added to the variance specification but rarely are in practice.

It is worth noting that in this model, the correlations ρ_{ij} between the mean equation disturbances are required to be fixed over time. Thus the covariances across the N series evolve as functions of the variances rather than independently of them. This formulation parallels the constant conditional correlation multivariate GARCH model of Bollerslev (1990) discussed above, and represents an important limitation of the model. It does, however, imply that MSV models are highly parsimonious, and the number of parameters scales directly with the number of variables in the system. For example, in the context of a bivariate MSV model, there are eight parameters to estimate.[2]

Harvey, Ruiz and Shephard (1994) propose estimating the model using quasimaximum likelihood (QML) via the Kalman filter. However, Danielsson (1998) argues that their QML approach results in inefficient estimation. An alternative approach to estimating MSV models is to make use of Bayesian Markov Chain Monte Carlo (MCMC) methods, as proposed by Jacquier, Polson and Rossi (1995).[3]

KEY CONCEPTS

The key terms to be able to define and explain from this chapter are

- non-linearity
- conditional variance
- maximum likelihood
- lagrange multiplier test
- asymmetry in volatility
- constant conditional correlation
- diagonal VECH
- news impact curve
- volatility clustering
- GARCH model
- Wald test
- likelihood ratio test
- GJR specification
- exponentially weighted moving average
- BEKK model
- GARCH-in-mean

Appendix 9.1 Parameter Estimation Using Maximum Likelihood

For simplicity, this appendix will consider by way of illustration the bivariate regression case with homoscedastic errors (i.e., assuming that there is no ARCH and that the variance of the errors is constant over time). Suppose that the linear regression model of interest is of the form

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad \textbf{(9A.1)}$$

Assuming that $u_t \sim \text{N}(0, \sigma^2)$, then $y_t \sim \text{N}(\beta_1 + \beta_2 x_t, \sigma^2)$ so that the probability density function for a normally distributed random variable with this mean and variance is given by

$$f(y_t \mid \beta_1 + \beta_2 x_t, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2}\right\} \tag{9A.2}$$

The probability density is a function of the data given the parameters. Successive values of y_t would trace out the familiar bell-shaped curve of the normal distribution. Since the ys are iid, the joint probability density function (pdf) for all the ys can be expressed as a product of the individual density functions

$$\begin{aligned} & f(y_1, y_2, \ldots, y_T \mid \beta_1 + \beta_2 x_1, \beta_1 + \beta_2 x_2, \ldots, \beta_1 + \beta_2 x_T, \sigma^2) \\ & = f(y_1 \mid \beta_1 + \beta_2 x_2, \sigma^2) f(y_2 \mid \beta_1 + \beta_2 x_2, \sigma^2) \ldots f(y_T \mid \beta_1 + \beta_2 x_T, \sigma^2) \\ & = \prod_{t=1}^{T} f(y_t \mid \beta_1 + \beta_2 x_t, \sigma^2) \quad \text{for } t = 1, \ldots, T \end{aligned} \tag{9A.3}$$

The term on the LHS of this expression is known as the *joint density* and the terms on the RHS are known as the *marginal densities*. This result follows from the independence of the y values, in the same way as under elementary probability, for three independent events A, B and C, the probability of A, B and C all happening is the probability of A multiplied by the probability of B multiplied by the probability of C. Equation (9A.3) shows the probability of obtaining all of the values of y that did occur. Substituting into equation (9A.3) for every y_t from (9A.2), and using the result that $Ae^{x_1} \times Ae^{x_2} \times \cdots Ae^{x_T} = A^T(e^{x_1} \times e^{x_2} \times \cdots \times e^{x_T}) = A^T e^{(x_1 + x_2 + \cdots + x_T)}$, the following expression is obtained

$$\begin{aligned} & f(y_1, y_2, \ldots, y_T \mid \beta_1 + \beta_2 x_t, \sigma^2) \\ & = \frac{1}{\sigma^T(\sqrt{2\pi})^T} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2}\right\} \end{aligned} \tag{9A.4}$$

This is the joint density of all of the ys given the values of x_t, β_1, β_2 and σ^2. However, the typical situation that occurs in practice is the reverse of the above situation – that is, the x_t and y_t are given and β_1, β_2, σ^2 are to be estimated. If this is the case, then $f(\bullet)$ is known as a likelihood function, denoted $LF(\beta_1, \beta_2, \sigma^2)$, which would be written

$$LF(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^T(\sqrt{2\pi})^T} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2}\right\} \tag{9A.5}$$

Maximum likelihood estimation involves choosing parameter values (β_1, β_2, σ^2) that maximise this function. Doing this ensures that the values of the parameters are chosen that maximise the likelihood that we would have actually observed the *y*s that we did. It is necessary to differentiate (9A.5) w.r.t. β_1, β_2, σ^2, but equation (9A.5) is a product containing *T* terms, and so would be difficult to differentiate.

Fortunately, since $\max_x f(x) = \max_x \ln(f(x))$, logs of equation (9A.3) can be taken, and the resulting expression differentiated, knowing that the same optimal values for the parameters will be chosen in both cases. Then, using the various laws for transforming functions containing logarithms, the log-likelihood function, *LLF* is obtained

$$LLF = -T\ln\sigma - \frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \tag{9A.6}$$

which is equivalent to

$$LLF = -\frac{T}{2}\ln\sigma^2 - \frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \tag{9A.7}$$

Only the first part of the RHS of equation (9A.6) has been changed in equation (9A.7) to make σ^2 appear in that part of the expression rather than σ.

Remembering the result that

$$\frac{\partial}{\partial x}(\ln(x)) = \frac{1}{x}$$

and differentiating equation (9A.7) w.r.t. β_1, β_2, σ^2, the following expressions for the first derivatives are obtained

$$\frac{\partial LLF}{\partial \beta_1} = -\frac{1}{2}\sum\frac{(y_t - \beta_1 - \beta_2 x_t).2. - 1}{\sigma^2} \tag{9A.8}$$

$$\frac{\partial LLF}{\partial \beta_2} = -\frac{1}{2}\sum\frac{(y_t - \beta_1 - \beta_2 x_t).2. - x_t}{\sigma^2} \tag{9A.9}$$

$$\frac{\partial LLF}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} + \frac{1}{2}\sum\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^4} \tag{9A.10}$$

Setting equations (9A.8)–(9A.10) to zero to minimise the functions, and placing hats above the parameters to denote the maximum likelihood estimators, from equation (9A.8)

$$\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) = 0 \tag{9A.11}$$

$$\sum y_t - \sum \hat{\beta}_1 - \sum \hat{\beta}_2 x_t = 0 \tag{9A.12}$$

$$\sum y_t - T\hat{\beta}_1 - \hat{\beta}_2 \sum x_t = 0 \tag{9A.13}$$

$$\frac{1}{T}\sum y_t - \hat{\beta}_1 - \hat{\beta}_2 \frac{1}{T}\sum x_t = 0 \tag{9A.14}$$

Recall that

$$\frac{1}{T}\sum y_t = \bar{y}_t$$

the mean of *y* and similarly for *x*, an estimator for $\hat{\beta}_1$ can finally be derived

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \tag{9A.15}$$

From equation (9A.9)

$$\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) x_t = 0 \tag{9A.16}$$

$$\sum y_t x_t - \sum \hat{\beta}_1 x_t - \sum \hat{\beta}_2 x_t^2 = 0 \tag{9A.17}$$

$$\sum y_t x_t - \hat{\beta}_1 \sum x_t - \hat{\beta}_2 \sum x_t^2 = 0 \tag{9A.18}$$

$$\hat{\beta}_2 \sum x_t^2 = \sum y_t x_t - (\bar{y} - \hat{\beta}_2 \bar{x}) \sum x_t \tag{9A.19}$$

$$\hat{\beta}_2 \sum x_t^2 = \sum y_t x_t - T\bar{x}\bar{y} + \hat{\beta}_2 T\bar{x}^2 \tag{9A.20}$$

$$\hat{\beta}_2 \left(\sum x_t^2 - T\bar{x}^2\right) = \sum y_t x_t - T\bar{x}\bar{y} \tag{9A.21}$$

$$\hat{\beta}_2 = \frac{\sum y_t x_t - T\bar{x}\bar{y}}{\left(\sum x_t^2 - T\bar{x}^2\right)} \tag{9A.22}$$

From equation (9A.10)

$$\frac{T}{\hat{\sigma}^2} = \frac{1}{\hat{\sigma}^4} \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2 \tag{9A.23}$$

Rearranging,

$$\hat{\sigma}^2 = \frac{1}{T} \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2 \tag{9A.24}$$

But the term in parentheses on the RHS of equation (9A.24) is the residual for time *t* (i.e., the actual minus the fitted value), so

$$\hat{\sigma}^2 = \frac{1}{T}\sum \hat{u}_t^2 \quad \textbf{(9A.25)}$$

How do these formulae compare with the OLS estimators? Equations (9A.15) and (9A.22) are identical to those of OLS. So maximum likelihood and OLS will deliver identical estimates of the intercept and slope coefficients. However, the estimate of $\hat{\sigma}^2$ in equation (9A.25) is different. The OLS estimator was

$$\hat{\sigma}^2 = \frac{1}{T-k}\sum \hat{u}_t^2 \quad \textbf{(9A.26)}$$

and it was also shown that the OLS estimator is unbiased. Therefore, the ML estimator of the error variance must be biased, although it is consistent, since as $T \rightarrow \infty$, $T - k \approx T$.

Note that the derivation above could also have been conducted using matrix rather than sigma algebra. The resulting estimators for the intercept and slope coefficients would still be identical to those of OLS, while the estimate of the error variance would again be biased. It is also worth noting that the ML estimator is consistent and asymptotically efficient. Derivation of the ML estimator for the GARCH *LLF* is algebraically difficult and therefore beyond the scope of this book.

SELF-STUDY QUESTIONS

1. (a) What stylised features of financial data cannot be explained using linear time series models?

(b) Which of these features could be modelled using a GARCH(1,1) process?

(c) Why, in recent empirical research, have researchers preferred GARCH(1,1) models to pure ARCH(*p*)?

(d) Describe two extensions to the original GARCH model. What additional characteristics of financial data might they be able to capture?

(e) Consider the following GARCH(1,1) model

$$y_t = \mu + u_t, \qquad u_t \sim \mathrm{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2$$

If y_t is a daily stock return series, what range of values are likely for the coefficients μ, α_0, α_1 and β?

(f) Suppose that a researcher wanted to test the null hypothesis that $\alpha_1 + \beta = 1$ in the equation for part (e). Explain how this might be achieved within the maximum likelihood framework.

(g) Suppose now that the researcher had estimated the above GARCH model for a series of returns on a stock index and obtained the following parameter estimates: $\hat{\mu} = 0.0023$, $\hat{\alpha}_0 = 0.0172$, $\hat{\beta} = 0.9811$, $\hat{\alpha}_1 = 0.1251$. If the researcher has data available up to and including time T, write down a set of equations in σ_t^2 and u_t^2 their lagged values, which could be employed to produce one-, two-, and three-step-ahead forecasts for the conditional variance of y_t.

(h) Suppose now that the coefficient estimate of $\hat{\beta}$ for this model is 0.98 instead. By reconsidering the forecast expressions you derived in part (g), explain what would happen to the forecasts in this case.

2. (a) Discuss briefly the principles behind maximum likelihood.

(b) Describe briefly the three hypothesis testing procedures that are available under maximum likelihood estimation. Which is likely to be the easiest to calculate in practice, and why?

(c) OLS and maximum likelihood are used to estimate the parameters of a standard linear regression model. Will they give the same estimates? Explain your answer.

3. (a) Distinguish between the terms ‘conditional variance’ and ‘unconditional variance’. Which of the two is more likely to be relevant for producing:

i. one-step-ahead volatility forecasts

ii. twenty-step-ahead volatility forecasts.

(b) If u_t follows a GARCH(1,1) process, what would be the likely result if a regression of the form in Question 1(e) were estimated using OLS and assuming a constant conditional variance?

(c) Compare and contrast the following models for volatility, noting their strengths and weaknesses:

i. Historical volatility

ii. EWMA

iii. GARCH(1,1)

iv. Implied volatility.

4. Suppose that a researcher is interested in modelling the correlation between the returns of the NYSE and LSE markets.

(a) Write down a simple diagonal *VECH* model for this problem. Discuss the values for the coefficient estimates that you would expect.

(b) Suppose that weekly correlation forecasts for two weeks ahead are required. Describe a procedure for constructing such forecasts from a set of daily returns data for the two market indices.

(c) What other approaches to correlation modelling are available?

(d) What are the strengths and weaknesses of multivariate GARCH models relative to the alternatives that you propose in part (c)?

5. (a) What is a news impact curve? Using a spreadsheet or otherwise, construct the news impact curve for the following estimated EGARCH and GARCH models, setting the lagged conditional variance to the value of the unconditional variance (estimated from the sample data rather than the mode parameter estimates), which is 0.096

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$

$$\ln\left(\sigma_t^2\right) = \alpha_0 + \alpha_1 \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha_2 \ln\left(\sigma_{t-1}^2\right) + \alpha_3 \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right]$$

	GARCH	EGARCH
μ	−0.0130	−0.0278
	(0.0669)	(0.0855)
α_0	0.0019	0.0823
	(0.0017)	(0.5728)
α_1	0.1022**	−0.0214
	(0.0333)	(0.0332)

α_2	0.9050^{**}	0.9639^{**}
	(0.0175)	(0.0136)
α_3	–	0.2326^{**}
		(0.0795)

(b) In fact, the models in part (a) were estimated using daily foreign exchange returns. How can financial theory explain the patterns observed in the news impact curves?

1 The *BEKK* acronym arises from the fact that early versions of the paper also listed Baba and Krafts as co-authors.

2 This compares with nine for a diagonal *VECH* MGARCH model and 21 for the unrestricted MGARCH.

3 See Chib and Greenberg (1996) for an extensive but very technical discussion of the intricacies of the MCMC technique.