



## Modelling volatility and correlation

---

### Learning Outcomes

In this chapter, you will learn how to

- Discuss the features of data that motivate the use of GARCH models
  - Explain how conditional volatility models are estimated
  - Test for ‘ARCH-effects’ in time series data
  - Produce forecasts from GARCH models
  - Contrast various models from the GARCH family
  - Discuss the three hypothesis testing procedures available under maximum likelihood estimation
  - Construct multivariate conditional volatility models and compare between alternative specifications
  - Estimate univariate and multivariate GARCH models in EViews
- 

### 8.1 Motivations: an excursion into non-linearity land

All of the models that have been discussed in chapters 2–7 of this book have been linear in nature – that is, the model is linear in the parameters, so that there is one parameter multiplied by each variable in the model. For example, a structural model could be something like

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u \quad (8.1)$$

or more compactly  $y = X\beta + u$ . It was additionally assumed that  $u_i \sim N(0, \sigma^2)$ .

The linear paradigm as described above is a useful one. The properties of linear estimators are very well researched and very well understood. Many models that appear, *prima facie*, to be non-linear, can be made linear

by taking logarithms or some other suitable transformation. However, it is likely that many relationships in finance are intrinsically non-linear. As Campbell, Lo and MacKinlay (1997) state, the payoffs to options are non-linear in some of the input variables, and investors' willingness to trade off returns and risks are also non-linear. These observations provide clear motivations for consideration of non-linear models in a variety of circumstances in order to capture better the relevant features of the data.

Linear structural (and time series) models such as (8.1) are also unable to explain a number of important features common to much financial data, including:

- *Leptokurtosis* – that is, the tendency for financial asset returns to have distributions that exhibit fat tails and excess peakedness at the mean.
- *Volatility clustering or volatility pooling* – the tendency for volatility in financial markets to appear in bunches. Thus large returns (of either sign) are expected to follow large returns, and small returns (of either sign) to follow small returns. A plausible explanation for this phenomenon, which seems to be an almost universal feature of asset return series in finance, is that the information arrivals which drive price changes themselves occur in bunches rather than being evenly spaced over time.
- *Leverage effects* – the tendency for volatility to rise more following a large price fall than following a price rise of the same magnitude.

Campbell, Lo and MacKinlay (1997) broadly define a non-linear data generating process as one where the current value of the series is related non-linearly to current and previous values of the error term

$$y_t = f(u_t, u_{t-1}, u_{t-2}, \dots) \quad (8.2)$$

where  $u_t$  is an iid error term and  $f$  is a non-linear function. According to Campbell, Lo and MacKinlay, a more workable and slightly more specific definition of a non-linear model is given by the equation

$$y_t = g(u_{t-1}, u_{t-2}, \dots) + u_t \sigma^2(u_{t-1}, u_{t-2}, \dots) \quad (8.3)$$

where  $g$  is a function of past error terms only, and  $\sigma^2$  can be interpreted as a variance term, since it is multiplied by the current value of the error. Campbell, Lo and MacKinlay usefully characterise models with non-linear  $g(\bullet)$  as being non-linear in mean, while those with non-linear  $\sigma(\bullet)^2$  are characterised as being non-linear in variance.

Models can be linear in mean and variance (e.g. the CLRM, ARMA models) or linear in mean, but non-linear in variance (e.g. GARCH models).

Models could also be classified as non-linear in mean but linear in variance (e.g. bicorrelations models, a simple example of which is of the following form (see Brooks and Hinich, 1999))

$$y_t = \alpha_0 + \alpha_1 y_{t-1} y_{t-2} + u_t \quad (8.4)$$

Finally, models can be non-linear in both mean and variance (e.g. the hybrid threshold model with GARCH errors employed by Brooks, 2001).

### 8.1.1 *Types of non-linear models*

There are an infinite number of different types of non-linear model. However, only a small number of non-linear models have been found to be useful for modelling financial data. The most popular non-linear financial models are the ARCH or GARCH models used for modelling and forecasting volatility, and switching models, which allow the behaviour of a series to follow different processes at different points in time. Models for volatility and correlation will be discussed in this chapter, with switching models being covered in chapter 9.

### 8.1.2 *Testing for non-linearity*

How can it be determined whether a non-linear model may potentially be appropriate for the data? The answer to this question should come at least in part from financial theory: a non-linear model should be used where financial theory suggests that the relationship between variables should be such as to require a non-linear model. But the linear versus non-linear choice may also be made partly on statistical grounds – deciding whether a linear specification is sufficient to describe all of the most important features of the data at hand.

So what tools are available to detect non-linear behaviour in financial time series? Unfortunately, ‘traditional’ tools of time series analysis (such as estimates of the autocorrelation or partial autocorrelation function, or ‘spectral analysis’, which involves looking at the data in the frequency domain) are likely to be of little use. Such tools may find no evidence of linear structure in the data, but this would not necessarily imply that the same observations are independent of one another.

However, there are a number of tests for non-linear patterns in time series that are available to the researcher. These tests can broadly be split into two types: general tests and specific tests. General tests, also sometimes called ‘portmanteau’ tests, are usually designed to detect many departures from randomness in data. The implication is that such tests will

detect a variety of non-linear structures in data, although these tests are unlikely to tell the researcher which type of non-linearity is present! Perhaps the simplest general test for non-linearity is Ramsey's RESET test discussed in chapter 4, although there are many other popular tests available. One of the most widely used tests is known as the BDS test (see Brock *et al.*, 1996) named after the three authors who first developed it. BDS is a pure hypothesis test. That is, it has as its null hypothesis that the data are pure noise (completely random), and it has been argued to have power to detect a variety of departures from randomness – linear or non-linear stochastic processes, deterministic chaos, etc. (see Brock *et al.*, 1991). The BDS test follows a standard normal distribution under the null hypothesis. The details of this test, and others, are technical and beyond the scope of this book, although computer code for BDS estimation is now widely available free of charge on the Internet.

As well as applying the BDS test to raw data in an attempt to 'see if there is anything there', another suggested use of the test is as a model diagnostic. The idea is that a proposed model (e.g. a linear model, GARCH, or some other non-linear model) is estimated, and the test applied to the (standardised) residuals in order to 'see what is left'. If the proposed model is adequate, the standardised residuals should be white noise, while if the postulated model is insufficient to capture all of the relevant features of the data, the BDS test statistic for the standardised residuals will be statistically significant. This is an excellent idea in theory, but has difficulties in practice. First, if the postulated model is a non-linear one (such as GARCH), the asymptotic distribution of the test statistic will be altered, so that it will no longer follow a normal distribution. This requires new critical values to be constructed via simulation for every type of non-linear model whose residuals are to be tested. More seriously, if a non-linear model is fitted to the data, any remaining structure is typically garbled, resulting in the test either being unable to detect additional structure present in the data (see Brooks and Henry, 2000) or selecting as adequate a model which is not even in the correct class for that data generating process (see Brooks and Heravi, 1999).

The BDS test is available in EViews. To run it on a given series, simply open the series to be tested (which may be a set of raw data or residuals from an estimated model) so that it appears as a spreadsheet. Then select the **View** menu and **BDS Independence Test...** You will then be offered various options. Further details are given in the *EViews User's Guide*.

Other popular tests for non-linear structure in time series data include the bispectrum test due to Hinich (1982), the bicorrelation test (see Hsieh,

1993; Hinich, 1996; or Brooks and Hinich, 1999 for its multivariate generalisation).

Most applications of the above tests conclude that there is non-linear dependence in financial asset returns series, but that the dependence is best characterised by a GARCH-type process (see Hinich and Patterson, 1985; Baillie and Bollerslev, 1989; Brooks, 1996; and the references therein for applications of non-linearity tests to financial data).

Specific tests, on the other hand, are usually designed to have power to find specific types of non-linear structure. Specific tests are unlikely to detect other forms of non-linearities in the data, but their results will by definition offer a class of models that should be relevant for the data at hand. Examples of specific tests will be offered later in this and subsequent chapters.

## **8.2 Models for volatility**

Modelling and forecasting stock market volatility has been the subject of vast empirical and theoretical investigation over the past decade or so by academics and practitioners alike. There are a number of motivations for this line of inquiry. Arguably, volatility is one of the most important concepts in the whole of finance. Volatility, as measured by the standard deviation or variance of returns, is often used as a crude measure of the total risk of financial assets. Many value-at-risk models for measuring market risk require the estimation or forecast of a volatility parameter. The volatility of stock market prices also enters directly into the Black-Scholes formula for deriving the prices of traded options.

The next few sections will discuss various models that are appropriate to capture the stylised features of volatility, discussed below, that have been observed in the literature.

## **8.3 Historical volatility**

The simplest model for volatility is the historical estimate. Historical volatility simply involves calculating the variance (or standard deviation) of returns in the usual way over some historical period, and this then becomes the volatility forecast for all future periods. The historical average variance (or standard deviation) was traditionally used as the volatility input to options pricing models, although there is a growing body of evidence suggesting that the use of volatility predicted from more

sophisticated time series models will lead to more accurate option valuations (see, for example, Akgiray, 1989; or Chu and Freund, 1996). Historical volatility is still useful as a benchmark for comparing the forecasting ability of more complex time models.

#### **8.4 Implied volatility models**

All pricing models for financial options require a volatility estimate or forecast as an input. Given the price of a traded option obtained from transactions data, it is possible to determine the volatility forecast over the lifetime of the option implied by the option's valuation. For example, if the standard Black–Scholes model is used, the option price, the time to maturity, a risk-free rate of interest, the strike price and the current value of the underlying asset, are all either specified in the details of the options contracts or are available from market data. Therefore, given all of these quantities, it is possible to use a numerical procedure, such as the method of bisections or Newton–Raphson to derive the volatility implied by the option (see Watsham and Parramore, 2004). This implied volatility is the market's forecast of the volatility of underlying asset returns over the lifetime of the option.

#### **8.5 Exponentially weighted moving average models**

The exponentially weighted moving average (EWMA) is essentially a simple extension of the historical average volatility measure, which allows more recent observations to have a stronger impact on the forecast of volatility than older data points. Under an EWMA specification, the latest observation carries the largest weight, and weights associated with previous observations decline exponentially over time. This approach has two advantages over the simple historical model. First, volatility is in practice likely to be affected more by recent events, which carry more weight, than events further in the past. Second, the effect on volatility of a single given observation declines at an exponential rate as weights attached to recent events fall. On the other hand, the simple historical approach could lead to an abrupt change in volatility once the shock falls out of the measurement sample. And if the shock is still included in a relatively long measurement sample period, then an abnormally large observation will imply that the forecast will remain at an artificially high level even if the market is subsequently tranquil. The exponentially weighted moving

average model can be expressed in several ways, e.g.

$$\sigma_t^2 = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j (r_{t-j} - \bar{r})^2 \quad (8.5)$$

where  $\sigma_t^2$  is the estimate of the variance for period  $t$ , which also becomes the forecast of future volatility for all periods,  $\bar{r}$  is the average return estimated over the observations and  $\lambda$  is the ‘decay factor’, which determines how much weight is given to recent versus older observations. The decay factor could be estimated, but in many studies is set at 0.94 as recommended by RiskMetrics, producers of popular risk measurement software. Note also that RiskMetrics and many academic papers assume that the average return,  $\bar{r}$ , is zero. For data that is of daily frequency or higher, this is not an unreasonable assumption, and is likely to lead to negligible loss of accuracy since it will typically be very small. Obviously, in practice, an infinite number of observations will not be available on the series, so that the sum in (8.5) must be truncated at some fixed lag. As with exponential smoothing models, the forecast from an EWMA model for all prediction horizons is the most recent weighted average estimate.

It is worth noting two important limitations of EWMA models. First, while there are several methods that could be used to compute the EWMA, the crucial element in each case is to remember that when the infinite sum in (8.5) is replaced with a finite sum of observable data, the weights from the given expression will now sum to less than one. In the case of small samples, this could make a large difference to the computed EWMA and thus a correction may be necessary. Second, most time-series models, such as GARCH (see below), will have forecasts that tend towards the unconditional variance of the series as the prediction horizon increases. This is a good property for a volatility forecasting model to have, since it is well known that volatility series are ‘mean-reverting’. This implies that if they are currently at a high level relative to their historic average, they will have a tendency to fall back towards their average level, while if they are at a low level relative to their historic average, they will have a tendency to rise back towards the average. This feature is accounted for in GARCH volatility forecasting models, but not by EWMA.

## 8.6 Autoregressive volatility models

Autoregressive volatility models are a relatively simple example from the class of stochastic volatility specifications. The idea is that a time series of observations on some volatility proxy are obtained. The standard

Box–Jenkins-type procedures for estimating autoregressive (or ARMA) models can then be applied to this series. If the quantity of interest in the study is a daily volatility estimate, two natural proxies have been employed in the literature: squared daily returns, or daily range estimators. Producing a series of daily squared returns trivially involves taking a column of observed returns and squaring each observation. The squared return at each point in time,  $t$ , then becomes the daily volatility estimate for day  $t$ . A range estimator typically involves calculating the log of the ratio of the highest observed price to the lowest observed price for trading day  $t$ , which then becomes the volatility estimate for day  $t$

$$\sigma_t^2 = \log \left( \frac{high_t}{low_t} \right) \quad (8.6)$$

Given either the squared daily return or the range estimator, a standard autoregressive model is estimated, with the coefficients  $\beta_i$  estimated using OLS (or maximum likelihood – see below). The forecasts are also produced in the usual fashion discussed in chapter 5 in the context of ARMA models

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \varepsilon_t \quad (8.7)$$

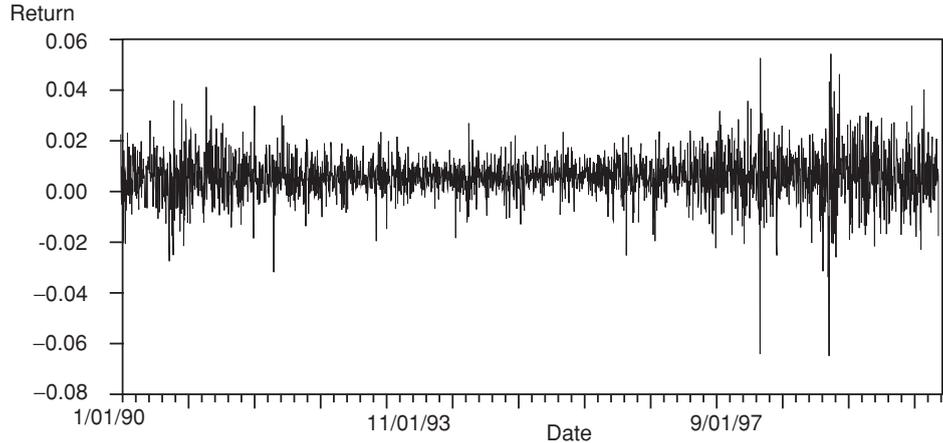
## 8.7 Autoregressive conditionally heteroscedastic (ARCH) models

One particular non-linear model in widespread usage in finance is known as an ‘ARCH’ model (ARCH stands for ‘autoregressive conditionally heteroscedastic’). To see why this class of models is useful, recall that a typical structural model could be expressed by an equation of the form given in (8.1) above with  $u_t \sim N(0, \sigma^2)$ . The assumption of the CLRM that the variance of the errors is constant is known as *homoscedasticity* (i.e. it is assumed that  $\text{var}(u_t) = \sigma^2$ ). If the variance of the errors is not constant, this would be known as *heteroscedasticity*. As was explained in chapter 4, if the errors are heteroscedastic, but assumed homoscedastic, an implication would be that standard error estimates could be wrong. It is unlikely in the context of financial time series that the variance of the errors will be constant over time, and hence it makes sense to consider a model that does not assume that the variance is constant, and which describes how the variance of the errors evolves.

Another important feature of many series of financial asset returns that provides a motivation for the ARCH class of models, is known as ‘volatility clustering’ or ‘volatility pooling’. Volatility clustering describes

**Figure 8.1**

Daily S&P returns  
for January  
1990–December  
1999



the tendency of large changes in asset prices (of either sign) to follow large changes and small changes (of either sign) to follow small changes. In other words, the current level of volatility tends to be positively correlated with its level during the immediately preceding periods. This phenomenon is demonstrated in figure 8.1, which plots daily S&P500 returns for January 1990–December 1999.

The important point to note from figure 8.1 is that *volatility occurs in bursts*. There appears to have been a prolonged period of relative tranquility in the market during the mid-1990s, evidenced by only relatively small positive and negative returns. On the other hand, during mid-1997 to late 1998, there was far more volatility, when many large positive and large negative returns were observed during a short space of time. Abusing the terminology slightly, it could be stated that ‘volatility is autocorrelated’.

How could this phenomenon, which is common to many series of financial asset returns, be parameterised (modelled)? One approach is to use an ARCH model. To understand how the model works, a definition of the conditional variance of a random variable,  $u_t$ , is required. The distinction between the conditional and unconditional variances of a random variable is exactly the same as that of the conditional and unconditional mean. The conditional variance of  $u_t$  may be denoted  $\sigma_t^2$ , which is written as

$$\sigma_t^2 = \text{var}(u_t | u_{t-1}, u_{t-2}, \dots) = E[(u_t - E(u_t))^2 | u_{t-1}, u_{t-2}, \dots] \quad (8.8)$$

It is usually assumed that  $E(u_t) = 0$ , so

$$\sigma_t^2 = \text{var}(u_t | u_{t-1}, u_{t-2}, \dots) = E[u_t^2 | u_{t-1}, u_{t-2}, \dots] \quad (8.9)$$

Equation (8.9) states that the conditional variance of a zero mean normally distributed random variable  $u_t$  is equal to the conditional expected

value of the square of  $u_t$ . Under the ARCH model, the ‘autocorrelation in volatility’ is modelled by allowing the conditional variance of the error term,  $\sigma_t^2$ , to depend on the immediately previous value of the squared error

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 \quad (8.10)$$

The above model is known as an ARCH(1), since the conditional variance depends on only one lagged squared error. Notice that (8.10) is only a partial model, since nothing has been said yet about the conditional mean. Under ARCH, the conditional mean equation (which describes how the dependent variable,  $y_t$ , varies over time) could take almost any form that the researcher wishes. One example of a full model would be

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad u_t \sim N(0, \sigma_t^2) \quad (8.11)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 \quad (8.12)$$

The model given by (8.11) and (8.12) could easily be extended to the general case where the error variance depends on  $q$  lags of squared errors, which would be known as an ARCH( $q$ ) model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 \quad (8.13)$$

Instead of calling the conditional variance  $\sigma_t^2$ , in the literature it is often called  $h_t$ , so that the model would be written

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad u_t \sim N(0, h_t) \quad (8.14)$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 \quad (8.15)$$

The remainder of this chapter will use  $\sigma_t^2$  to denote the conditional variance at time  $t$ , except for computer instructions where  $h_t$  will be used since it is easier not to use Greek letters.

### 8.7.1 Another way of expressing ARCH models

For illustration, consider an ARCH(1). The model can be expressed in two ways that look different but are in fact identical. The first is as given in (8.11) and (8.12) above. The second way would be as follows

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad (8.16)$$

$$u_t = v_t \sigma_t \quad v_t \sim N(0, 1) \quad (8.17)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 \quad (8.18)$$

The form of the model given in (8.11) and (8.12) is more commonly presented, although specifying the model as in (8.16)–(8.18) is required in

order to use a GARCH process in a simulation study (see chapter 12). To show that the two methods for expressing the model are equivalent, consider that in (8.17),  $v_t$  is normally distributed with zero mean and unit variance, so that  $u_t$  will also be normally distributed with zero mean and variance  $\sigma_t^2$ .

### 8.7.2 Non-negativity constraints

Since  $h_t$  is a conditional variance, its value must always be strictly positive; a negative variance at any point in time would be meaningless. The variables on the RHS of the conditional variance equation are all squares of lagged errors, and so by definition will not be negative. In order to ensure that these always result in positive conditional variance estimates, all of the coefficients in the conditional variance are usually required to be non-negative. If one or more of the coefficients were to take on a negative value, then for a sufficiently large lagged squared innovation term attached to that coefficient, the fitted value from the model for the conditional variance could be negative. This would clearly be nonsensical. So, for example, in the case of (8.18), the non-negativity condition would be  $\alpha_0 \geq 0$  and  $\alpha_1 \geq 0$ . More generally, for an ARCH( $q$ ) model, all coefficients would be required to be non-negative:  $\alpha_i \geq 0 \forall i = 0, 1, 2, \dots, q$ . In fact, this is a sufficient but not necessary condition for non-negativity of the conditional variance (i.e. it is a slightly stronger condition than is actually necessary).

### 8.7.3 Testing for 'ARCH effects'

A test for determining whether 'ARCH-effects' are present in the residuals of an estimated model may be conducted using the steps outlined in box 8.1.

Thus, the test is one of a joint null hypothesis that all  $q$  lags of the squared residuals have coefficient values that are not significantly different from zero. If the value of the test statistic is greater than the critical value from the  $\chi^2$  distribution, then reject the null hypothesis. The test can also be thought of as a test for autocorrelation in the squared residuals. As well as testing the residuals of an estimated model, the ARCH test is frequently applied to raw returns data.

### 8.7.4 Testing for 'ARCH effects' in exchange rate returns using EViews

Before estimating a GARCH-type model, it is sensible first to compute the Engle (1982) test for ARCH effects to make sure that this class of models is appropriate for the data. This exercise (and the remaining exercises of this chapter), will employ returns on the daily exchange rates where there are

**Box 8.1** Testing for 'ARCH effects'

- (1) Run any postulated linear regression of the form given in the equation above, e.g.

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad (8.19)$$

saving the residuals,  $\hat{u}_t$ .

- (2) Square the residuals, and regress them on
- $q$
- own lags to test for ARCH of order
- $q$
- , i.e. run the regression

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2 + \dots + \gamma_q \hat{u}_{t-q}^2 + v_t \quad (8.20)$$

where  $v_t$  is an error term.Obtain  $R^2$  from this regression.

- (3) The test statistic is defined as
- $TR^2$
- (the number of observations multiplied by the coefficient of multiple correlation) from the last regression, and is distributed as a
- $\chi^2(q)$
- .
- 
- (4) The null and alternative hypotheses are

$$H_0 : \gamma_1 = 0 \text{ and } \gamma_2 = 0 \text{ and } \gamma_3 = 0 \text{ and } \dots \text{ and } \gamma_q = 0$$

$$H_1 : \gamma_1 \neq 0 \text{ or } \gamma_2 \neq 0 \text{ or } \gamma_3 \neq 0 \text{ or } \dots \text{ or } \gamma_q \neq 0$$

1,827 observations. Models of this kind are inevitably more data intensive than those based on simple linear regressions, and hence, everything else being equal, they work better when the data are sampled daily rather than at a lower frequency.

A test for the presence of ARCH in the residuals is calculated by regressing the squared residuals on a constant and  $p$  lags, where  $p$  is set by the user. As an example, assume that  $p$  is set to 5. The first step is to estimate a linear model so that the residuals can be tested for ARCH. From the main menu, select **Quick** and then select **Estimate Equation**. In the Equation Specification Editor, input **rgbp c ar(1) ma(1)** which will estimate an ARMA(1,1) for the pound-dollar returns.<sup>1</sup> Select the **Least Squares (NLA and ARMA)** procedure to estimate the model, using the whole sample period and press the **OK** button (output not shown).

The next step is to click on **View** from the Equation Window and to select **Residual Tests** and then **Heteroskedasticity Tests...** In the 'Test type' box, choose **ARCH** and the number of lags to include is **5**, and press **OK**. The output below shows the Engle test results.

<sup>1</sup> Note that the (1,1) order has been chosen entirely arbitrarily at this stage. However, it is important to give some thought to the type and order of model used even if it is not of direct interest in the problem at hand (which will later be termed the 'conditional mean' equation), since the variance is measured around the mean and therefore any mis-specification in the mean is likely to lead to a mis-specified variance.

## Heteroskedasticity Test: ARCH

F-statistic	5.909063	Prob. F(5,1814)	0.0000
Obs*R-squared	29.16797	Prob. Chi-Square(5)	0.0000

## Test Equation:

Dependent Variable: RESID<sup>2</sup>

Method: Least Squares

Date: 09/06/07 Time: 14:41

Sample (adjusted): 7/14/2002 7/07/2007

Included observations: 1820 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.154689	0.011369	13.60633	0.0000
RESID <sup>2</sup> (-1)	0.118068	0.023475	5.029627	0.0000
RESID <sup>2</sup> (-2)	-0.006579	0.023625	-0.278463	0.7807
RESID <sup>2</sup> (-3)	0.029000	0.023617	1.227920	0.2196
RESID <sup>2</sup> (-4)	-0.032744	0.023623	-1.386086	0.1659
RESID <sup>2</sup> (-5)	-0.020316	0.023438	-0.866798	0.3862
R-squared	0.016026	Mean dependent var		0.169496
Adjusted R-squared	0.013314	S.D. dependent var		0.344448
S.E. of regression	0.342147	Akaike info criterion		0.696140
Sum squared resid	212.3554	Schwarz criterion		0.714293
Log likelihood	-627.4872	Hannan-Quinn criter.		0.702837
F-statistic	5.909063	Durbin-Watson stat		1.995904
Prob(F-statistic)	0.000020			

Both the  $F$ -version and the  $LM$ -statistic are very significant, suggesting the presence of ARCH in the pound-dollar returns.

### 8.7.5 Limitations of ARCH( $q$ ) models

ARCH provided a framework for the analysis and development of time series models of volatility. However, ARCH models themselves have rarely been used in the last decade or more, since they bring with them a number of difficulties:

- How should the value of  $q$ , the number of lags of the squared residual in the model, be decided? One approach to this problem would be the use of a likelihood ratio test, discussed later in this chapter, although there is no clearly best approach.
- The value of  $q$ , the number of lags of the squared error that are required to capture all of the dependence in the conditional variance, might be very large. This would result in a large conditional variance model that was not parsimonious. Engle (1982) circumvented this problem by

specifying an arbitrary linearly declining lag length on an ARCH(4)

$$\sigma_t^2 = \gamma_0 + \gamma_1(0.4\hat{u}_{t-1}^2 + 0.3\hat{u}_{t-2}^2 + 0.2\hat{u}_{t-3}^2 + 0.1\hat{u}_{t-4}^2) \quad (8.21)$$

such that only two parameters are required in the conditional variance equation ( $\gamma_0$  and  $\gamma_1$ ), rather than the five which would be required for an unrestricted ARCH(4).

- *Non-negativity constraints might be violated.* Everything else equal, the more parameters there are in the conditional variance equation, the more likely it is that one or more of them will have negative estimated values.

A natural extension of an ARCH( $q$ ) model which overcomes some of these problems is a GARCH model. In contrast with ARCH, GARCH models are extremely widely employed in practice.

## 8.8 Generalised ARCH (GARCH) models

The GARCH model was developed independently by Bollerslev (1986) and Taylor (1986). The GARCH model allows the conditional variance to be dependent upon previous own lags, so that the conditional variance equation in the simplest case is now

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (8.22)$$

This is a GARCH(1,1) model.  $\sigma_t^2$  is known as the *conditional variance* since it is a one-period ahead estimate for the variance calculated based on any past information thought relevant. Using the GARCH model it is possible to interpret the current fitted variance,  $h_t$ , as a weighted function of a long-term average value (dependent on  $\alpha_0$ ), information about volatility during the previous period ( $\alpha_1 u_{t-1}^2$ ) and the fitted variance from the model during the previous period ( $\beta \sigma_{t-1}^2$ ). Note that the GARCH model can be expressed in a form that shows that it is effectively an ARMA model for the conditional variance. To see this, consider that the squared return at time  $t$  relative to the conditional variance is given by

$$\varepsilon_t = u_t^2 - \sigma_t^2 \quad (8.23)$$

or

$$\sigma_t^2 = u_t^2 - \varepsilon_t \quad (8.24)$$

Using the latter expression to substitute in for the conditional variance in (8.22)

$$u_t^2 - \varepsilon_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta(u_{t-1}^2 - \varepsilon_{t-1}) \quad (8.25)$$

Rearranging

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta u_{t-1}^2 - \beta \varepsilon_{t-1} + \varepsilon_t \quad (8.26)$$

so that

$$u_t^2 = \alpha_0 + (\alpha_1 + \beta) u_{t-1}^2 - \beta \varepsilon_{t-1} + \varepsilon_t \quad (8.27)$$

This final expression is an ARMA(1,1) process for the squared errors.

Why is GARCH a better and therefore a far more widely used model than ARCH? The answer is that the former is more parsimonious, and avoids overfitting. Consequently, the model is less likely to breach non-negativity constraints. In order to illustrate why the model is parsimonious, first take the conditional variance equation in the GARCH(1,1) case, subtract 1 from each of the time subscripts of the conditional variance equation in (8.22), so that the following expression would be obtained

$$\sigma_{t-1}^2 = \alpha_0 + \alpha_1 u_{t-2}^2 + \beta \sigma_{t-2}^2 \quad (8.28)$$

and subtracting 1 from each of the time subscripts again

$$\sigma_{t-2}^2 = \alpha_0 + \alpha_1 u_{t-3}^2 + \beta \sigma_{t-3}^2 \quad (8.29)$$

Substituting into (8.22) for  $\sigma_{t-1}^2$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta (\alpha_0 + \alpha_1 u_{t-2}^2 + \beta \sigma_{t-2}^2) \quad (8.30)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0 \beta + \alpha_1 \beta u_{t-2}^2 + \beta^2 \sigma_{t-2}^2 \quad (8.31)$$

Now substituting into (8.31) for  $\sigma_{t-2}^2$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0 \beta + \alpha_1 \beta u_{t-2}^2 + \beta^2 (\alpha_0 + \alpha_1 u_{t-3}^2 + \beta \sigma_{t-3}^2) \quad (8.32)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0 \beta + \alpha_1 \beta u_{t-2}^2 + \alpha_0 \beta^2 + \alpha_1 \beta^2 u_{t-3}^2 + \beta^3 \sigma_{t-3}^2 \quad (8.33)$$

$$\sigma_t^2 = \alpha_0 (1 + \beta + \beta^2) + \alpha_1 u_{t-1}^2 (1 + \beta L + \beta^2 L^2) + \beta^3 \sigma_{t-3}^2 \quad (8.34)$$

An infinite number of successive substitutions of this kind would yield

$$\sigma_t^2 = \alpha_0 (1 + \beta + \beta^2 + \dots) + \alpha_1 u_{t-1}^2 (1 + \beta L + \beta^2 L^2 + \dots) + \beta^\infty \sigma_0^2 \quad (8.35)$$

The first expression on the RHS of (8.35) is simply a constant, and as the number of observations tends to infinity,  $\beta^\infty$  will tend to zero. Hence, the GARCH(1,1) model can be written as

$$\sigma_t^2 = \gamma_0 + \alpha_1 u_{t-1}^2 (1 + \beta L + \beta^2 L^2 + \dots) \quad (8.36)$$

$$= \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + \dots, \quad (8.37)$$

which is a restricted infinite order ARCH model. Thus the GARCH(1,1) model, containing only three parameters in the conditional variance

equation, is a very parsimonious model, that allows an infinite number of past squared errors to influence the current conditional variance.

The GARCH(1,1) model can be extended to a GARCH( $p,q$ ) formulation, where the current conditional variance is parameterised to depend upon  $q$  lags of the squared error and  $p$  lags of the conditional variance

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 \\ &\quad + \beta_2 \sigma_{t-2}^2 + \cdots + \beta_p \sigma_{t-p}^2\end{aligned}\quad (8.38)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (8.39)$$

But in general a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data, and rarely is any higher order model estimated or even entertained in the academic finance literature.

### 8.8.1 The unconditional variance under a GARCH specification

The conditional variance is changing, but the unconditional variance of  $u_t$  is constant and given by

$$\text{var}(u_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta)} \quad (8.40)$$

so long as  $\alpha_1 + \beta < 1$ . For  $\alpha_1 + \beta \geq 1$ , the unconditional variance of  $u_t$  is not defined, and this would be termed ‘non-stationarity in variance’.  $\alpha_1 + \beta = 1$  would be known as a ‘unit root in variance’, also termed ‘Integrated GARCH’ or IGARCH. Non-stationarity in variance does not have a strong theoretical motivation for its existence, as would be the case for non-stationarity in the mean (e.g. of a price series). Furthermore, a GARCH model whose coefficients imply non-stationarity in variance would have some highly undesirable properties. One illustration of these relates to the forecasts of variance made from such models. For stationary GARCH models, conditional variance forecasts converge upon the long-term average value of the variance as the prediction horizon increases (see below). For IGARCH processes, this convergence will not happen, while for  $\alpha_1 + \beta > 1$ , the conditional variance forecast will tend to infinity as the forecast horizon increases!

## 8.9 Estimation of ARCH/GARCH models

Since the model is no longer of the usual linear form, OLS cannot be used for GARCH model estimation. There are a variety of reasons for this, but the simplest and most fundamental is that OLS minimises the residual

**Box 8.2** Estimating an ARCH or GARCH model

- (1) Specify the appropriate equations for the mean and the variance – e.g. an AR(1)-GARCH(1,1) model

$$y_t = \mu + \phi y_{t-1} + u_t, u_t \sim N(0, \sigma_t^2) \quad (8.41)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (8.42)$$

- (2) Specify the log-likelihood function (*LLF*) to maximise under a normality assumption for the disturbances

$$L = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^T (y_t - \mu - \phi y_{t-1})^2 / \sigma_t^2 \quad (8.43)$$

- (3) The computer will maximise the function and generate parameter values that maximise the *LLF* and will construct their standard errors.

sum of squares. The RSS depends only on the parameters in the conditional mean equation, and not the conditional variance, and hence RSS minimisation is no longer an appropriate objective.

In order to estimate models from the GARCH family, another technique known as *maximum likelihood* is employed. Essentially, the method works by finding the most likely values of the parameters given the actual data. More specifically, a log-likelihood function is formed and the values of the parameters that maximise it are sought. Maximum likelihood estimation can be employed to find parameter values for both linear and non-linear models. The steps involved in actually estimating an ARCH or GARCH model are shown in box 8.2.

The following section will elaborate on points 2 and 3 above, explaining how the *LLF* is derived.

### 8.9.1 Parameter estimation using maximum likelihood

As stated above, under maximum likelihood estimation, a set of parameter values are chosen that are most likely to have produced the observed data. This is done by first forming a *likelihood function*, denoted *LF*. *LF* will be a multiplicative function of the actual data, which will consequently be difficult to maximise with respect to the parameters. Therefore, its logarithm is taken in order to turn *LF* into an additive function of the sample data, i.e. the *LLF*. A derivation of the maximum likelihood (ML) estimator in the context of the simple bivariate regression model with homoscedasticity is given in the appendix to this chapter. Essentially, deriving the ML estimators involves differentiating the *LLF* with respect to the parameters. But how does this help in estimating heteroscedastic models? How can the

method outlined in the appendix for homoscedastic models be modified for application to GARCH model estimation?

In the context of conditional heteroscedasticity models, the model is  $y_t = \mu + \phi y_{t-1} + u_t$ ,  $u_t \sim N(0, \sigma_t^2)$ , so that the variance of the errors has been modified from being assumed constant,  $\sigma^2$ , to being time-varying,  $\sigma_t^2$ , with the equation for the conditional variance as previously. The *LLF* relevant for a GARCH model can be constructed in the same way as for the homoscedastic case by replacing

$$\frac{T}{2} \log \sigma^2$$

with the equivalent for time-varying variance

$$\frac{1}{2} \sum_{t=1}^T \log \sigma_t^2$$

and replacing  $\sigma^2$  in the denominator of the last part of the expression with  $\sigma_t^2$  (see the appendix to this chapter). Derivation of this result from first principles is beyond the scope of this text, but the log-likelihood function for the above model with time-varying conditional variance and normally distributed errors is given by (8.43) in box 8.2.

Intuitively, maximising the *LLF* involves jointly minimising

$$\sum_{t=1}^T \log \sigma_t^2$$

and

$$\sum_{t=1}^T \frac{(y_t - \mu - \phi y_{t-1})^2}{\sigma_t^2}$$

(since these terms appear preceded with a negative sign in the *LLF*, and

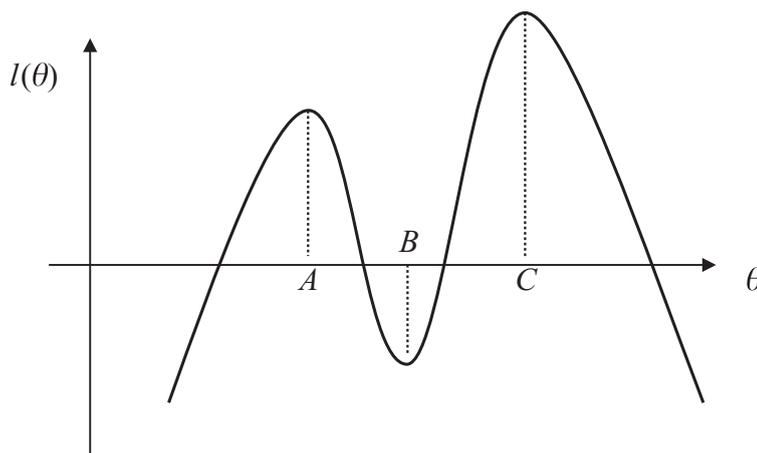
$$-\frac{T}{2} \log(2\pi)$$

is just a constant with respect to the parameters). Minimising these terms jointly also implies minimising the error variance, as described in chapter 3. Unfortunately, maximising the *LLF* for a model with time-varying variances is trickier than in the homoscedastic case. Analytical derivatives of the *LLF* in (8.43) with respect to the parameters have been developed, but only in the context of the simplest examples of GARCH specifications. Moreover, the resulting formulae are complex, so a numerical procedure is often used instead to maximise the log-likelihood function.

Essentially, all methods work by ‘searching’ over the parameter-space until the values of the parameters that maximise the log-likelihood

**Figure 8.2**

The problem of local optima in maximum likelihood estimation



function are found. EViews employs an iterative technique for maximising the *LLF*. This means that, given a set of initial guesses for the parameter estimates, these parameter values are updated at each iteration until the program determines that an optimum has been reached. If the *LLF* has only one maximum with respect to the parameter values, any optimisation method should be able to find it – although some methods will take longer than others. A detailed presentation of the various methods available is beyond the scope of this book. However, as is often the case with non-linear models such as GARCH, the *LLF* can have many local maxima, so that different algorithms could find different local maxima of the *LLF*. Hence readers should be warned that different optimisation procedures could lead to different coefficient estimates and especially different estimates of the standard errors (see Brooks, Burke and Persaud, 2001 or 2003 for details). In such instances, a good set of initial parameter guesses is essential.

Local optima or multimodalities in the likelihood surface present potentially serious drawbacks with the maximum likelihood approach to estimating the parameters of a GARCH model, as shown in figure 8.2.

Suppose that the model contains only one parameter,  $\theta$ , so that the log-likelihood function is to be maximised with respect to this one parameter. In figure 8.2, the value of the *LLF* for each value of  $\theta$  is denoted  $l(\theta)$ . Clearly,  $l(\theta)$  reaches a global maximum when  $\theta = C$ , and a local maximum when  $\theta = A$ . This demonstrates the importance of good initial guesses for the parameters. Any initial guesses to the left of  $B$  are likely to lead to the selection of  $A$  rather than  $C$ . The situation is likely to be even worse in practice, since the log-likelihood function will be maximised with respect to several parameters, rather than one, and there could be

**Box 8.3** Using maximum likelihood estimation in practice

- (1) Set up the *LLF*.
- (2) Use regression to get *initial estimates* for the mean parameters.
- (3) Choose some initial guesses for the *conditional variance parameters*. In most software packages, the default initial values for the conditional variance parameters would be zero. This is unfortunate since zero parameter values often yield a local maximum of the likelihood function. So if possible, set plausible initial values away from zero.
- (4) Specify a *convergence criterion* – either by criterion or by value. When ‘by criterion’ is selected, the package will continue to search for ‘better’ parameter values that give a higher value of the *LLF* until the change in the value of the *LLF* between iterations is less than the specified convergence criterion. Choosing ‘by value’ will lead to the software searching until the change in the coefficient estimates are small enough. The default convergence criterion for EViews is 0.001, which means that convergence is achieved and the program will stop searching if the biggest percentage change in any of the coefficient estimates for the most recent iteration is smaller than 0.1%.

many local optima. Another possibility that would make optimisation difficult is when the *LLF* is flat around the maximum. So, for example, if the peak corresponding to  $C$  in figure 8.2, were flat rather than sharp, a range of values for  $\theta$  could lead to very similar values for the *LLF*, making it difficult to choose between them.

So, to explain again in more detail, the optimisation is done in the way shown in box 8.3.

The optimisation methods employed by EViews are based on the determination of the first and second derivatives of the log-likelihood function with respect to the parameter values at each iteration, known as the gradient and Hessian (the matrix of second derivatives of the *LLF* w.r.t the parameters), respectively. An algorithm for optimisation due to Berndt, Hall, Hall and Hausman (1974), known as BHHH, is available in EViews. BHHH employs only first derivatives (calculated numerically rather than analytically) and approximations to the second derivatives are calculated. Not calculating the actual Hessian at each iteration at each time step increases computational speed, but the approximation may be poor when the *LLF* is a long way from its maximum value, requiring more iterations to reach the optimum. The Marquardt algorithm, available in EViews, is a modification of BHHH (both of which are variants on the Gauss–Newton method) that incorporates a ‘correction’, the effect of which is to push the coefficient estimates more quickly to their optimal values. All of these optimisation methods are described in detail in Press *et al.* (1992).

### 8.9.2 Non-normality and maximum likelihood

Recall that the conditional normality assumption for  $u_t$  is essential in specifying the likelihood function. It is possible to test for non-normality using the following representation

$$u_t = v_t \sigma_t, v_t \sim N(0, 1) \quad (8.44)$$

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2} \quad (8.45)$$

Note that one would not expect  $u_t$  to be normally distributed – it is a  $N(0, \sigma_t^2)$  disturbance term from the regression model, which will imply it is likely to have fat tails. A plausible method to test for normality would be to construct the statistic

$$v_t = \frac{u_t}{\sigma_t} \quad (8.46)$$

which would be the model disturbance at each point in time  $t$  divided by the conditional standard deviation at that point in time. Thus, it is the  $v_t$  that are assumed to be normally distributed, not  $u_t$ . The sample counterpart would be

$$\hat{v}_t = \frac{\hat{u}_t}{\hat{\sigma}_t} \quad (8.47)$$

which is known as a standardised residual. Whether the  $\hat{v}_t$  are normal can be examined using any standard normality test, such as the Bera-Jarque. Typically,  $\hat{v}_t$  are still found to be leptokurtic, although less so than the  $\hat{u}_t$ . The upshot is that the GARCH model is able to capture some, although not all, of the leptokurtosis in the unconditional distribution of asset returns.

Is it a problem if  $\hat{v}_t$  are not normally distributed? Well, the answer is ‘not really’. Even if the conditional normality assumption does not hold, the parameter estimates will still be consistent if the equations for the mean and variance are correctly specified. However, in the context of non-normality, the usual standard error estimates will be inappropriate, and a different variance-covariance matrix estimator that is robust to non-normality, due to Bollerslev and Wooldridge (1992), should be used. This procedure (i.e. maximum likelihood with Bollerslev-Wooldridge standard errors) is known as *quasi-maximum likelihood*, or QML.

### 8.9.3 Estimating GARCH models in EViews

To estimate a GARCH-type model, open the equation specification dialog by selecting **Quick/Estimate Equation** or by selecting **Object/New Object/Equation...** Select ARCH from the ‘Estimation Settings’ selection box. The window in screenshot 8.1 will open.

**Screenshot 8.1**

Estimating a  
GARCH-type model

**Equation Estimation**

Specification Options

Mean equation  
Dependent followed by regressors and ARMA terms OR explicit

ARCH-M: None

Variance and distribution specification

Model: GARCH/TARCH

Variance regressors:

Order:  
ARCH: 1 Threshold order: 0

GARCH: 1

Restrictions: None

Error distribution: Normal (Gaussian)

Estimation settings

Method: ARCH - Autoregressive Conditional Heteroskedasticity

Sample: 7/07/2002 7/07/2007

OK Cancel

It is necessary to specify both the mean and the variance equations, as well as the estimation technique and sample.

#### The mean equation

The specification of the mean equation should be entered in the dependent variable edit box. Enter the specification by listing the dependent variable followed by the regressors. The constant term 'C' should also be included. If your specification includes an ARCH-M term (see later in this chapter), you should click on the appropriate button in the upper RHS of the dialog box to select the conditional standard deviation, the conditional variance, or the log of the conditional variance.

#### The variance equation

The edit box labelled 'Variance regressors' is where variables that are to be included in the variance specification should be listed. Note that EViews will always include a constant in the conditional variance, so that it is not necessary to add 'C' to the variance regressor list. Similarly, it is not

necessary to include the ARCH or GARCH terms in this box as they will be dealt with in other parts of the dialog box. Instead, enter here any exogenous variables or dummies that you wish to include in the conditional variance equation, or (as is usually the case), just leave this box blank.

#### Variance and distribution specification

Under the 'Variance and distribution Specification' label, choose the number of ARCH and GARCH terms. The default is to estimate with one ARCH and one GARCH term (i.e. one lag of the squared errors and one lag of the conditional variance, respectively). To estimate the standard GARCH model, leave the default 'GARCH/TARCH'. The other entries in this box describe more complicated variants of the standard GARCH specification, which are described in later sections of this chapter.

#### Estimation options

EViews provides a number of optional estimation settings. Clicking on the Options tab gives the options in screenshot 8.2 to be filled out as required.

#### Screenshot 8.2

GARCH model  
estimation options

The Heteroskedasticity Consistent Covariance option is used to compute the quasi-maximum likelihood (QML) covariances and standard errors using the methods described by Bollerslev and Wooldridge (1992). This option should be used if you suspect that the residuals are not conditionally normally distributed. Note that the parameter estimates will be (virtually) unchanged if this option is selected; only the estimated covariance matrix will be altered.

The log-likelihood functions for ARCH models are often not well behaved so that convergence may not be achieved with the default estimation settings. It is possible in EViews to select the iterative algorithm (Marquardt, BHHH/Gauss Newton), to change starting values, to increase the maximum number of iterations or to adjust the convergence criteria. For example, if convergence is not achieved, or implausible parameter estimates are obtained, it is sensible to re-do the estimation using a different set of starting values and/or a different optimisation algorithm.

Once the model has been estimated, EViews provides a variety of pieces of information and procedures for inference and diagnostic checking. For example, the following options are available on the View button:

- Actual, Fitted, Residual  
The residuals are displayed in various forms, such as table, graphs and standardised residuals.
- GARCH graph  
This graph plots the one-step ahead standard deviation,  $\sigma_t$ , or the conditional variance,  $\sigma_t^2$  for each observation in the sample.
- Covariance Matrix
- Coefficient Tests
- Residual Tests/Correlogram-Q statistics
- Residual Tests/Correlogram Squared Residuals
- Residual Tests/Histogram-Normality Test
- Residual Tests/ARCH LM Test.

#### ARCH model procedures

These options are all available by pressing the 'Proc' button following the estimation of a GARCH-type model:

- Make Residual Series
- Make GARCH Variance Series
- Forecast.

Estimating the GARCH(1,1) model for the yen-dollar ('rjpy') series using the instructions as listed above, and the default settings elsewhere would yield the results:

Dependent Variable: RJPY  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 09/06/07 Time: 18:02  
 Sample (adjusted): 7/08/2002 7/07/2007  
 Included observations: 1826 after adjustments  
 Convergence achieved after 10 iterations  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(2) + C(3)\*RESID(-1)^2 + C(4)\*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.005518	0.009396	0.587333	0.5570
Variance Equation				
C	0.001345	0.000526	2.558748	0.0105
RESID(-1)^2	0.028436	0.004108	6.922465	0.0000
GARCH(-1)	0.964139	0.005528	174.3976	0.0000
R-squared	-0.000091	Mean dependent var		0.001328
Adjusted R-squared	-0.001738	S.D. dependent var		0.439632
S.E. of regression	0.440014	Akaike info criterion		1.139389
Sum squared resid	352.7611	Schwarz criterion		1.151459
Log likelihood	-1036.262	Hannan-Quinn criter.		1.143841
Durbin-Watson stat	1.981759			

The coefficients on both the lagged squared residual and lagged conditional variance terms in the conditional variance equation are highly statistically significant. Also, as is typical of GARCH model estimates for financial asset returns data, the sum of the coefficients on the lagged squared error and lagged conditional variance is very close to unity (approximately 0.99). This implies that shocks to the conditional variance will be highly persistent. This can be seen by considering the equations for forecasting future values of the conditional variance using a GARCH model given in a subsequent section. A large sum of these coefficients will imply that a large positive or a large negative return will lead future forecasts of the variance to be high for a protracted period. The individual conditional variance coefficients are also as one would expect. The variance intercept term 'C' is very small, and the 'ARCH parameter' is around

0.03 while the coefficient on the lagged conditional variance ('GARCH') is larger at 0.96.

### 8.10 Extensions to the basic GARCH model

Since the GARCH model was developed, a huge number of extensions and variants have been proposed. A couple of the most important examples will be highlighted here. Interested readers who wish to investigate further are directed to a comprehensive survey by Bollerslev *et al.* (1992).

Many of the extensions to the GARCH model have been suggested as a consequence of perceived problems with standard GARCH( $p, q$ ) models. First, the non-negativity conditions may be violated by the estimated model. The only way to avoid this for sure would be to place artificial constraints on the model coefficients in order to force them to be non-negative. Second, GARCH models cannot account for leverage effects (explained below), although they can account for volatility clustering and leptokurtosis in a series. Finally, the model does not allow for any direct feedback between the conditional variance and the conditional mean.

Some of the most widely used and influential modifications to the model will now be examined. These may remove some of the restrictions or limitations of the basic model.

### 8.11 Asymmetric GARCH models

One of the primary restrictions of GARCH models is that they enforce a symmetric response of volatility to positive and negative shocks. This arises since the conditional variance in equations such as (8.39) is a function of the magnitudes of the lagged residuals and not their signs (in other words, by squaring the lagged error in (8.39), the sign is lost). However, it has been argued that a negative shock to financial time series is likely to cause volatility to rise by more than a positive shock of the same magnitude. In the case of equity returns, such asymmetries are typically attributed to *leverage effects*, whereby a fall in the value of a firm's stock causes the firm's debt to equity ratio to rise. This leads shareholders, who bear the residual risk of the firm, to perceive their future cashflow stream as being relatively more risky.

An alternative view is provided by the 'volatility-feedback' hypothesis. Assuming constant dividends, if expected returns increase when stock

price volatility increases, then stock prices should fall when volatility rises. Although asymmetries in returns series other than equities cannot be attributed to changing leverage, there is equally no reason to suppose that such asymmetries only exist in equity returns.

Two popular asymmetric formulations are explained below: the GJR model, named after the authors Glosten, Jagannathan and Runkle (1993), and the exponential GARCH (EGARCH) model proposed by Nelson (1991).

## 8.12 The GJR model

The GJR model is a simple extension of GARCH with an additional term added to account for possible asymmetries. The conditional variance is now given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma u_{t-1}^2 I_{t-1} \quad (8.48)$$

where  $I_{t-1} = 1$  if  $u_{t-1} < 0$   
 $= 0$  otherwise

For a leverage effect, we would see  $\gamma > 0$ . Notice now that the condition for non-negativity will be  $\alpha_0 > 0$ ,  $\alpha_1 > 0$ ,  $\beta \geq 0$ , and  $\alpha_1 + \gamma \geq 0$ . That is, the model is still admissible, even if  $\gamma < 0$ , provided that  $\alpha_1 + \gamma \geq 0$ .

### Example 8.1

To offer an illustration of the GJR approach, using monthly S&P500 returns from December 1979 until June 1998, the following results would be obtained, with  $t$ -ratios in parentheses

$$y_t = 0.172 \quad (3.198) \quad (8.49)$$

$$\sigma_t^2 = 1.243 + 0.015u_{t-1}^2 + 0.498\sigma_{t-1}^2 + 0.604u_{t-1}^2 I_{t-1} \quad (8.50)$$

(16.372) (0.437) (14.999) (5.772)

Note that the asymmetry term,  $\gamma$ , has the correct sign and is significant. To see how volatility rises more after a large negative shock than a large positive one, suppose that  $\sigma_{t-1}^2 = 0.823$ , and consider  $\hat{u}_{t-1} = \pm 0.5$ . If  $\hat{u}_{t-1} = 0.5$ , this implies that  $\sigma_t^2 = 1.65$ . However, a shock of the same magnitude but of opposite sign,  $\hat{u}_{t-1} = -0.5$ , implies that the fitted conditional variance for time  $t$  will be  $\sigma_t^2 = 1.80$ .

---

### 8.13 The EGARCH model

The exponential GARCH model was proposed by Nelson (1991). There are various ways to express the conditional variance equation, but one possible specification is given by

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \quad (8.51)$$

The model has several advantages over the pure GARCH specification. First, since the  $\log(\sigma_t^2)$  is modelled, then even if the parameters are negative,  $\sigma_t^2$  will be positive. There is thus no need to artificially impose non-negativity constraints on the model parameters. Second, asymmetries are allowed for under the EGARCH formulation, since if the relationship between volatility and returns is negative,  $\gamma$ , will be negative.

Note that in the original formulation, Nelson assumed a Generalised Error Distribution (GED) structure for the errors. GED is a very broad family of distributions that can be used for many types of series. However, owing to its computational ease and intuitive interpretation, almost all applications of EGARCH employ conditionally normal errors as discussed above rather than using GED.

### 8.14 GJR and EGARCH in EViews

The main menu screen for GARCH estimation demonstrates that a number of variants on the standard GARCH model are available. Arguably most important of these are asymmetric models, such as the TGARCH ('threshold' GARCH), which is also known as the GJR model, and the EGARCH model. To estimate a GJR model in EViews, from the GARCH model equation specification screen (screenshot 8.1 above), change the 'Threshold Order' number from 0 to 1. To estimate an EGARCH model, change the 'GARCH/TARCH' model estimation default to 'EGARCH'.

Coefficient estimates for each of these specifications using the daily Japanese yen-US dollar returns data are given in the next two output tables, respectively. For both specifications, the asymmetry terms ('(RESID<0)\* ARCH(1)' in the GJR model and 'RESID(-1)/@SQRT(GARCH(-1))') are not statistically significant (although it is almost significant in the case of the EGARCH model). Also in both cases, the coefficient estimates are negative, suggesting that positive shocks imply a higher next period conditional variance than negative shocks of the same sign.

Dependent Variable: RJPY  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 09/06/07 Time: 18:20  
 Sample (adjusted): 7/08/2002 7/07/2007  
 Included observations: 1826 after adjustments  
 Convergence achieved after 9 iterations  
 Presample variance: backcast (parameter = 0.7)  

$$\text{GARCH} = C(2) + C(3)*\text{RESID}(-1)^2 + C(4)*\text{RESID}(-1)^2*(\text{RESID}(-1)<0) + C(5)*\text{GARCH}(-1)$$

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.005588	0.009602	0.581934	0.5606
Variance Equation				
C	0.001361	0.000544	2.503534	0.0123
RESID(-1)^2	0.029036	0.005373	5.404209	0.0000
RESID(-1)^2*(RESID(-1)<0)	-0.001027	0.006140	-0.167301	0.8671
GARCH(-1)	0.963989	0.005644	170.7852	0.0000
R-squared	-0.000094	Mean dependent var		0.001328
Adjusted R-squared	-0.002291	S.D. dependent var		0.439632
S.E. of regression	0.440135	Akaike info criterion		1.140477
Sum squared resid	352.7622	Schwarz criterion		1.155564
Log likelihood	-1036.256	Hannan-Quinn criter.		1.146042
Durbin-Watson stat	1.981753			

Dependent Variable: RJPY  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 09/06/07 Time: 18:18  
 Sample (adjusted): 7/08/2002 7/07/2007  
 Included observations: 1826 after adjustments  
 Convergence achieved after 12 iterations  
 Presample variance: backcast (parameter = 0.7)  

$$\text{LOG}(\text{GARCH}) = C(2) + C(3)*\text{ABS}(\text{RESID}(-1)/\text{SQRT}(\text{GARCH}(-1))) + C(4)*\text{RESID}(-1)/\text{SQRT}(\text{GARCH}(-1)) + C(5)*\text{LOG}(\text{GARCH}(-1))$$

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.003756	0.010025	0.374722	0.7079
Variance Equation				
C(2)	-1.262782	0.194243	-6.501047	0.0000
C(3)	0.214215	0.034226	6.258919	0.0000
C(4)	-0.046461	0.024983	-1.859751	0.0629
C(5)	0.329164	0.112572	2.924037	0.0035
R-squared	-0.000031	Mean dependent var		0.001328
Adjusted R-squared	-0.002227	S.D. dependent var		0.439632
S.E. of regression	0.440121	Akaike info criterion		1.183216
Sum squared resid	352.7398	Schwarz criterion		1.198303
Log likelihood	-1075.276	Hannan-Quinn criter.		1.188781
Durbin-Watson stat	1.981879			

This is the opposite to what would have been expected in the case of the application of a GARCH model to a set of stock returns. But arguably, neither the *leverage effect* or *volatility feedback* explanations for asymmetries in the context of stocks apply here. For a positive return shock, this implies more yen per dollar and therefore a strengthening dollar and a weakening yen. Thus the results suggest that a strengthening dollar (weakening yen) leads to higher next period volatility than when the yen strengthens by the same amount.

### 8.15 Tests for asymmetries in volatility

Engle and Ng (1993) have proposed a set of tests for asymmetry in volatility, known as sign and size bias tests. The Engle and Ng tests should thus be used to determine whether an asymmetric model is required for a given series, or whether the symmetric GARCH model can be deemed adequate. In practice, the Engle–Ng tests are usually applied to the residuals of a GARCH fit to the returns data. Define  $S_{t-1}^-$  as an indicator dummy that takes the value 1 if  $\hat{u}_{t-1} < 0$  and zero otherwise. The test for sign bias is based on the significance or otherwise of  $\phi_1$  in

$$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + v_t \quad (8.52)$$

where  $v_t$  is an iid error term. If positive and negative shocks to  $\hat{u}_{t-1}$  impact differently upon the conditional variance, then  $\phi_1$  will be statistically significant.

It could also be the case that the magnitude or size of the shock will affect whether the response of volatility to shocks is symmetric or not. In this case, a negative size bias test would be conducted, based on a regression where  $S_{t-1}^-$  is now used as a slope dummy variable. Negative size bias is argued to be present if  $\phi_1$  is statistically significant in the regression

$$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- u_{t-1} + v_t \quad (8.53)$$

Finally, defining  $S_{t-1}^+ = 1 - S_{t-1}^-$ , so that  $S_{t-1}^+$  picks out the observations with positive innovations, Engle and Ng propose a joint test for sign and size bias based on the regression

$$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + \phi_2 S_{t-1}^- u_{t-1} + \phi_3 S_{t-1}^+ u_{t-1} + v_t \quad (8.54)$$

Significance of  $\phi_1$  indicates the presence of sign bias, where positive and negative shocks have differing impacts upon future volatility, compared with the symmetric response required by the standard GARCH

formulation. On the other hand, the significance of  $\phi_2$  or  $\phi_3$  would suggest the presence of size bias, where not only the sign but the magnitude of the shock is important. A joint test statistic is formulated in the standard fashion by calculating  $TR^2$  from regression (8.54), which will asymptotically follow a  $\chi^2$  distribution with 3 degrees of freedom under the null hypothesis of no asymmetric effects.

### 8.15.1 News impact curves

A pictorial representation of the degree of asymmetry of volatility to positive and negative shocks is given by the news impact curve introduced by Pagan and Schwert (1990). The news impact curve plots the next-period volatility ( $\sigma_t^2$ ) that would arise from various positive and negative values of  $u_{t-1}$ , given an estimated model. The curves are drawn by using the estimated conditional variance equation for the model under consideration, with its given coefficient estimates, and with the lagged conditional variance set to the unconditional variance. Then, successive values of  $u_{t-1}$  are used in the equation to determine what the corresponding values of  $\sigma_t^2$  derived from the model would be. For example, consider the GARCH and GJR model estimates given above for the S&P500 data from EViews. Values of  $u_{t-1}$  in the range  $(-1, +1)$  are substituted into the equations in each case to investigate the impact on the conditional variance during the next period. The resulting news impact curves for the GARCH and GJR models are given in figure 8.3.

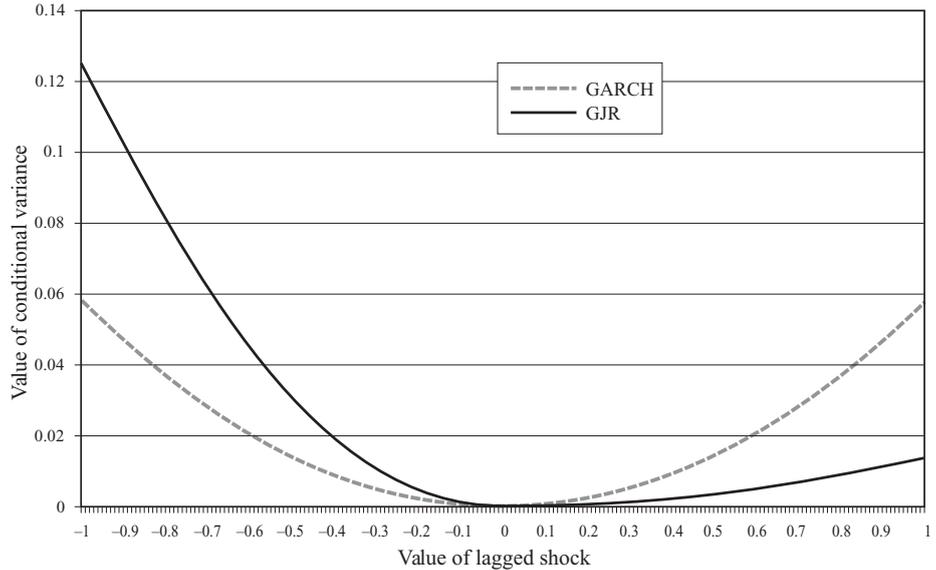
As can be seen from figure 8.3, the GARCH news impact curve (the grey line) is of course symmetrical about zero, so that a shock of given magnitude will have the same impact on the future conditional variance whatever its sign. On the other hand, the GJR news impact curve (the black line) is asymmetric, with negative shocks having more impact on future volatility than positive shocks of the same magnitude. It can also be seen that a negative shock of given magnitude will have a bigger impact under GJR than would be implied by a GARCH model, while a positive shock of given magnitude will have more impact under GARCH than GJR. The latter result arises as a result of the reduction in the value of  $\alpha_1$ , the coefficient on the lagged squared error, when the asymmetry term is included in the model.

## 8.16 GARCH-in-mean

Most models used in finance suppose that investors should be rewarded for taking additional risk by obtaining a higher return. One way to

**Figure 8.3**

News impact curves for S&P500 returns using coefficients implied from GARCH and GJR model estimates



operationalise this concept is to let the return of a security be partly determined by its risk. Engle, Lilien and Robins (1987) suggested an ARCH-M specification, where the conditional variance of asset returns enters into the conditional mean equation. Since GARCH models are now considerably more popular than ARCH, it is more common to estimate a GARCH-M model. An example of a GARCH-M model is given by the specification

$$y_t = \mu + \delta\sigma_{t-1} + u_t, u_t \sim N(0, \sigma_t^2) \quad (8.55)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (8.56)$$

If  $\delta$  is positive and statistically significant, then increased risk, given by an increase in the conditional variance, leads to a rise in the mean return. Thus  $\delta$  can be interpreted as a risk premium. In some empirical applications, the conditional variance term,  $\sigma_{t-1}^2$ , appears directly in the conditional mean equation, rather than in square root form,  $\sigma_{t-1}$ . Also, in some applications the term is contemporaneous,  $\sigma_t^2$ , rather than lagged.

### 8.16.1 GARCH-M estimation in EViews

The GARCH-M model with the conditional standard deviation term in the mean, estimated using the rjpy data in EViews from the main GARCH menu as described above, would give the following results:

Dependent Variable: RJPY  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 09/06/07 Time: 18:58  
 Sample (adjusted): 7/08/2002 7/07/2007  
 Included observations: 1826 after adjustments  
 Convergence achieved after 18 iterations  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(3) + C(4)\*RESID(-1)^2 + C(5)\*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
SQRT(GARCH)	-0.068943	0.124958	-0.551729	0.5811
C	0.033279	0.051802	0.642436	0.5206
Variance Equation				
C	0.001373	0.000529	2.594929	0.0095
RESID(-1)^2	0.028886	0.004150	6.960374	0.0000
GARCH(-1)	0.963568	0.005580	172.6828	0.0000
R-squared	0.000034	Mean dependent var		0.001328
Adjusted R-squared	-0.002162	S.D. dependent var		0.439632
S.E. of regression	0.440107	Akaike info criterion		1.140302
Sum squared resid	352.7170	Schwarz criterion		1.155390
Log likelihood	-1036.096	Hannan-Quinn criter.		1.145867
F-statistic	0.015541	Durbin-Watson stat		1.982106
Prob(F-statistic)	0.999526			

In this case, the estimated parameter on the mean equation has a negative sign but is not statistically significant. We would thus conclude that for these currency returns, there is no feedback from the conditional variance to the conditional mean.

## 8.17 Uses of GARCH-type models including volatility forecasting

Essentially GARCH models are useful because they can be used to model the volatility of a series over time. It is possible to combine together more than one of the time series models that have been considered so far in this book, to obtain more complex 'hybrid' models. Such models can account for a number of important features of financial series at the same time - e.g. an ARMA-EGARCH(1,1)-M model; the potential complexity of the model is limited only by the imagination!

GARCH-type models can be used to forecast volatility. GARCH is a model to describe movements in the conditional variance of an error term,

$u_t$ , which may not appear particularly useful. But it is possible to show that

$$\text{var}(y_t | y_{t-1}, y_{t-2}, \dots) = \text{var}(u_t | u_{t-1}, u_{t-2}, \dots) \quad (8.57)$$

So the conditional variance of  $y$ , given its previous values, is the same as the conditional variance of  $u$ , given its previous values. Hence, modelling  $\sigma_t^2$  will give models and forecasts for the variance of  $y_t$  as well. Thus, if the dependent variable in a regression,  $y_t$  is an asset return series, forecasts of  $\sigma_t^2$  will be forecasts of the future variance of  $y_t$ . So one primary usage of GARCH-type models is in forecasting volatility. This can be useful in, for example, the pricing of financial options where volatility is an input to the pricing model. For example, the value of a ‘plain vanilla’ call option is a function of the current value of the underlying, the strike price, the time to maturity, the risk free interest rate and volatility. The required volatility, to obtain an appropriate options price, is really the volatility of the underlying asset expected over the lifetime of the option. As stated previously, it is possible to use a simple historical average measure as the forecast of future volatility, but another method that seems more appropriate would be to use a time series model such as GARCH to compute the volatility forecasts. The forecasting ability of various models is considered in a paper by Day and Lewis (1992), discussed in detail below.

Producing forecasts from models of the GARCH class is relatively simple, and the algebra involved is very similar to that required to obtain forecasts from ARMA models. An illustration is given by example 8.2.

### Example 8.2

Consider the following GARCH(1,1) model

$$y_t = \mu + u_t, u_t \sim N(0, \sigma_t^2) \quad (8.58)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (8.59)$$

Suppose that the researcher had estimated the above GARCH model for a series of returns on a stock index and obtained the following parameter estimates:  $\hat{\mu} = 0.0023$ ,  $\hat{\alpha}_0 = 0.0172$ ,  $\hat{\beta} = 0.7811$ ,  $\hat{\alpha}_1 = 0.1251$ . If the researcher has data available up to and including time  $T$ , write down a set of equations in  $\sigma_t^2$  and  $u_t^2$  and their lagged values, which could be employed to produce one-, two-, and three-step-ahead forecasts for the conditional variance of  $y_t$ .

What is needed is to generate forecasts of  $\sigma_{T+1}^2 | \Omega_T$ ,  $\sigma_{T+2}^2 | \Omega_T$ , ...,  $\sigma_{T+s}^2 | \Omega_T$  where  $\Omega_T$  denotes all information available up to and including

observation  $T$ . For time  $T$ , the conditional variance equation is given by (8.59). Adding one to each of the time subscripts of this equation, and then two, and then three would yield equations (8.60)–(8.62)

$$\sigma_{T+1}^2 = \alpha_0 + \alpha_1 u_T^2 + \beta \sigma_T^2 \quad (8.60)$$

$$\sigma_{T+2}^2 = \alpha_0 + \alpha_1 u_{T+1}^2 + \beta \sigma_{T+1}^2 \quad (8.61)$$

$$\sigma_{T+3}^2 = \alpha_0 + \alpha_1 u_{T+2}^2 + \beta \sigma_{T+2}^2 \quad (8.62)$$

Let  $\sigma_{1,T}^{f^2}$  be the one-step-ahead forecast for  $\sigma^2$  made at time  $T$ . This is easy to calculate since, at time  $T$ , the values of all the terms on the RHS are known.  $\sigma_{1,T}^{f^2}$  would be obtained by taking the conditional expectation of (8.60).

Given  $\sigma_{1,T}^{f^2}$ , how is  $\sigma_{2,T}^{f^2}$ , the two-step-ahead forecast for  $\sigma^2$  made at time  $T$ , calculated?

$$\sigma_{1,T}^{f^2} = \alpha_0 + \alpha_1 u_T^2 + \beta \sigma_T^2 \quad (8.63)$$

From (8.61), it is possible to write

$$\sigma_{2,T}^{f^2} = \alpha_0 + \alpha_1 E(u_{T+1}^2 | \Omega_T) + \beta \sigma_{1,T}^{f^2} \quad (8.64)$$

where  $E(u_{T+1}^2 | \Omega_T)$  is the expectation, made at time  $T$ , of  $u_{T+1}^2$ , which is the squared disturbance term. It is necessary to find  $E(u_{T+1}^2 | \Omega_T)$ , using the expression for the variance of a random variable  $u_t$ . The model assumes that the series  $u_t$  has zero mean, so that the variance can be written

$$\text{var}(u_t) = E[(u_t - E(u_t))^2] = E(u_t^2). \quad (8.65)$$

The conditional variance of  $u_t$  is  $\sigma_t^2$ , so

$$\sigma_t^2 | \Omega_t = E(u_t^2) \quad (8.66)$$

Turning this argument around, and applying it to the problem at hand

$$E(u_{T+1} | \Omega_t)^2 = \sigma_{T+1}^2 \quad (8.67)$$

but  $\sigma_{T+1}^2$  is not known at time  $T$ , so it is replaced with the forecast for it,  $\sigma_{1,T}^{f^2}$ , so that (8.64) becomes

$$\sigma_{2,T}^{f^2} = \alpha_0 + \alpha_1 \sigma_{1,T}^{f^2} + \beta \sigma_{1,T}^{f^2} \quad (8.68)$$

$$\sigma_{2,T}^{f^2} = \alpha_0 + (\alpha_1 + \beta) \sigma_{1,T}^{f^2} \quad (8.69)$$

What about the three-step-ahead forecast?

By similar arguments,

$$\sigma_{3,T}^{f^2} = E_T(\alpha_0 + \alpha_1 u_{T+2}^2 + \beta \sigma_{T+2}^2) \quad (8.70)$$

$$\sigma_{3,T}^{f^2} = \alpha_0 + (\alpha_1 + \beta) \sigma_{2,T}^{f^2} \quad (8.71)$$

$$\sigma_{3,T}^{f^2} = \alpha_0 + (\alpha_1 + \beta) [\alpha_0 + (\alpha_1 + \beta) \sigma_{1,T}^{f^2}] \quad (8.72)$$

$$\sigma_{3,T}^{f^2} = \alpha_0 + \alpha_0(\alpha_1 + \beta) + (\alpha_1 + \beta)^2 \sigma_{1,T}^{f^2} \quad (8.73)$$

Any  $s$ -step-ahead forecasts would be produced by

$$\sigma_{s,T}^{f^2} = \alpha_0 \sum_{i=1}^{s-1} (\alpha_1 + \beta)^{i-1} + (\alpha_1 + \beta)^{s-1} \sigma_{1,T}^{f^2} \quad (8.74)$$

for any value of  $s \geq 2$ .

It is worth noting at this point that variances, and therefore variance forecasts, are additive over time. This is a very useful property. Suppose, for example, that using daily foreign exchange returns, one-, two-, three-, four-, and five-step-ahead variance forecasts have been produced, i.e. a forecast has been constructed for each day of the next trading week. The forecasted variance for the whole week would simply be the sum of the five daily variance forecasts. If the standard deviation is the required volatility estimate rather than the variance, simply take the square root of the variance forecasts. Note also, however, that standard deviations are not additive. Hence, if daily standard deviations are the required volatility measure, they must be squared to turn them to variances. Then the variances would be added and the square root taken to obtain a weekly standard deviation.

### 8.17.1 Forecasting from GARCH models with EViews

Forecasts from any of the GARCH models that can be estimated using EViews are obtained by using only a sub-sample of available data for model estimation, and then by clicking on the 'Forecast' button that appears after the estimation of the required model has been completed. Suppose, for example, we stopped the estimation of the GARCH(1,1) model for the Japanese yen returns on 6 July 2005 so as to keep the last two years of data for forecasting (i.e. the 'Forecast sample' is 7/07/2005 7/07/2007). Then click **Proc/Forecast...** and the dialog box in screenshot 8.3 will then appear.

Again, several options are available, including providing a name for the conditional mean and for the conditional variance forecasts, or whether to produce static (a series of rolling single-step-ahead) or dynamic (multiple-step-ahead) forecasts. The dynamic and static forecast plots that would be produced are given in screenshots 8.4 and 8.5.

**Screenshot 8.3**

Forecasting from  
GARCH models

**Forecast** [X]

Forecast of  
Equation: GARCH11EQN      Series: RJPY

Series names  
Forecast name:   
S.E. (optional):   
GARCH(optional):

Method  
 Dynamic forecast  
 Static forecast  
 Structural (ignore ARMA)  
 Coef uncertainty in S.E. calc

Forecast sample

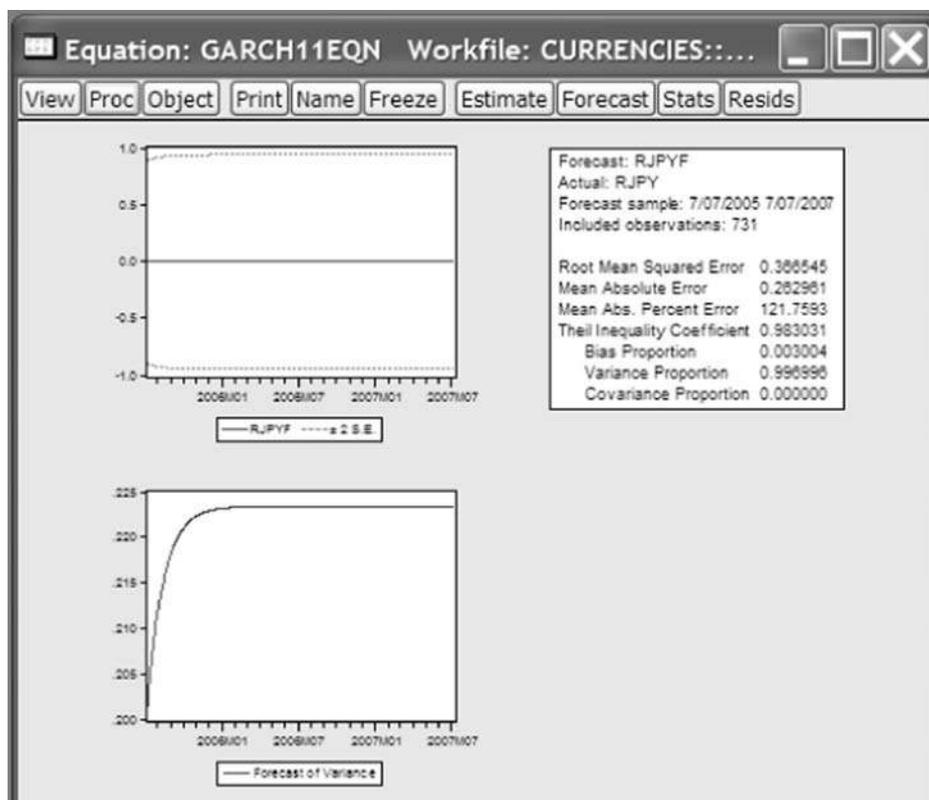
Output  
 Forecast graph  
 Forecast evaluation

Insert actuals for out-of-sample observations

OK      Cancel

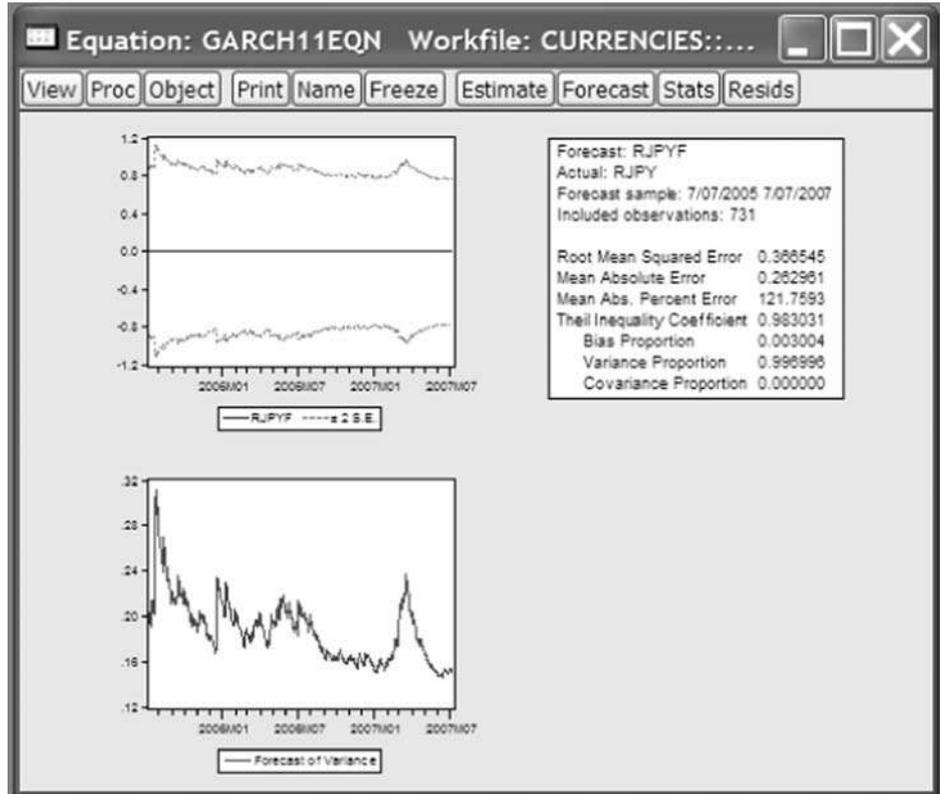
**Screenshot 8.4**

Dynamic forecasts  
of the conditional  
variance



**Screenshot 8.5**

Static forecasts of the conditional variance



### GARCH(1,1) Dynamic forecasts (2 years ahead)

The dynamic forecasts show a completely flat forecast structure for the mean (since the conditional mean equation includes only a constant term), while at the end of the in-sample estimation period, the value of the conditional variance was at a historically low level relative to its unconditional average. Therefore, the forecasts converge upon their long-term mean value from below as the forecast horizon increases. Notice also that there are no  $\pm 2$ -standard error band confidence intervals for the conditional variance forecasts; to compute these would require some kind of estimate of the variance of variance, which is beyond the scope of this book (and beyond the capability of the built-in functions of the EViews software). The conditional variance forecasts provide the basis for the standard error bands that are given by the dotted red lines around the conditional mean forecast. Because the conditional variance forecasts rise gradually as the forecast horizon increases, so the standard error bands widen slightly. The forecast evaluation statistics that are presented in the box to the right of the graphs are for the conditional mean forecasts.

GARCH(1,1) Static forecasts (1 month ahead – 22 days)

It is evident that the variance forecasts gradually fall over the out-of-sample period, although since these are a series of rolling one-step ahead forecasts for the conditional variance, they show much more volatility than for the dynamic forecasts. This volatility also results in more variability in the standard error bars around the conditional mean forecasts.

Predictions can be similarly produced for any member of the GARCH family that is estimable with the software.

### 8.18 Testing non-linear restrictions or testing hypotheses about non-linear models

The usual  $t$ - and  $F$ -tests are still valid in the context of non-linear models, but they are not flexible enough. For example, suppose that it is of interest to test a hypothesis that  $\alpha_1\beta = 1$ . Now that the model class has been extended to non-linear models, there is no reason to suppose that relevant restrictions are only linear.

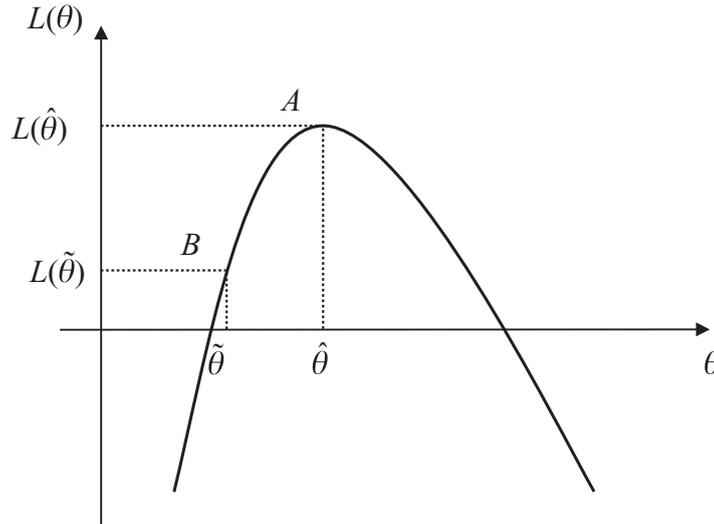
Under OLS estimation, the  $F$ -test procedure works by examining the degree to which the  $RSS$  rises when the restrictions are imposed. In very general terms, hypothesis testing under ML works in a similar fashion – that is, the procedure works by examining the degree to which the maximal value of the  $LLF$  falls upon imposing the restriction. If the  $LLF$  falls ‘a lot’, it would be concluded that the restrictions are not supported by the data and thus the hypothesis should be rejected.

There are three hypothesis testing procedures based on maximum likelihood principles: Wald, Likelihood ratio and Lagrange Multiplier. To illustrate briefly how each of these operates, consider a single parameter,  $\theta$  to be estimated, and denote the  $ML$  estimate as  $\hat{\theta}$  and a restricted estimate as  $\tilde{\theta}$ . Denoting the maximised value of the  $LLF$  by unconstrained  $ML$  as  $L(\hat{\theta})$  and the constrained optimum as  $L(\tilde{\theta})$ , the three testing procedures can be illustrated as in figure 8.4.

The tests all require the measurement of the ‘distance’ between the points  $A$  (representing the unconstrained maximised value of the log likelihood function) and  $B$  (representing the constrained value). The vertical distance forms the basis of the  $LR$  test. Twice this vertical distance is given by  $2[L(\hat{\theta}) - L(\tilde{\theta})] = 2\ln[l(\hat{\theta})/l(\tilde{\theta})]$ , where  $L$  denotes the log-likelihood function, and  $l$  denotes the likelihood function. The Wald test is based on the horizontal distance between  $\hat{\theta}$  and  $\tilde{\theta}$ , while the  $LM$  test compares the slopes of the curve at  $A$  and  $B$ . At  $A$ , the unrestricted maximum of the log-likelihood function, the slope of the curve is zero. But is it ‘significantly

**Figure 8.4**

Three approaches to hypothesis testing under maximum likelihood



steep' at  $L(\tilde{\theta})$ , i.e. at point  $B$ ? The steeper the curve is at  $B$ , the less likely the restriction is to be supported by the data.

Expressions for *LM* test statistics involve the first and second derivatives of the log-likelihood function with respect to the parameters at the constrained estimate. The first derivatives of the log-likelihood function are collectively known as the score vector, measuring the slope of the *LLF* for each possible value of the parameters. The expected values of the second derivatives comprise the information matrix, measuring the peakedness of the *LLF*, and how much higher the *LLF* value is at the optimum than in other places. This matrix of second derivatives is also used to construct the coefficient standard errors. The *LM* test involves estimating only a restricted regression, since the slope of the *LLF* at the maximum will be zero by definition. Since the restricted regression is usually easier to estimate than the unrestricted case, *LM* tests are usually the easiest of the three procedures to employ in practice. The reason that restricted regressions are usually simpler is that imposing the restrictions often means that some components in the model will be set to zero or combined under the null hypothesis, so that there are fewer parameters to estimate. The Wald test involves estimating only an unrestricted regression, and the usual OLS *t*-tests and *F*-tests are examples of Wald tests (since again, only unrestricted estimation occurs).

Of the three approaches to hypothesis testing in the maximum-likelihood framework, the likelihood ratio test is the most intuitively appealing, and therefore a deeper examination of it will be the subject of the following section; see Ghosh (1991, section 10.3) for further details.

### 8.18.1 Likelihood ratio tests

Likelihood ratio (*LR*) tests involve estimation under the null hypothesis and under the alternative, so that two models are estimated: an unrestricted model and a model where the restrictions have been imposed. The maximised values of the *LLF* for the restricted and unrestricted cases are ‘compared’. Suppose that the unconstrained model has been estimated and that a given maximised value of the *LLF*, denoted  $L_u$ , has been achieved. Suppose also that the model has been estimated imposing the constraint(s) and a new value of the *LLF* obtained, denoted  $L_r$ . The *LR* test statistic asymptotically follows a Chi-squared distribution and is given by

$$LR = -2(L_r - L_u) \sim \chi^2(m) \quad (8.75)$$

where  $m$  = number of restrictions. Note that the maximised value of the log-likelihood function will always be at least as big for the unrestricted model as for the restricted model, so that  $L_r \leq L_u$ . This rule is intuitive and comparable to the effect of imposing a restriction on a linear model estimated by OLS, that  $RRSS \geq URSS$ . Similarly, the equality between  $L_r$  and  $L_u$  will hold only when the restriction was already present in the data. Note, however, that the usual *F*-test is in fact a Wald test, and not a *LR* test – that is, it can be calculated using an unrestricted model only. The *F*-test approach based on comparing *RSS* arises conveniently as a result of the OLS algebra.

#### Example 8.3

A GARCH model is estimated and a maximised *LLF* of 66.85 is obtained. Suppose that a researcher wishes to test whether  $\beta = 0$  in (8.77)

$$y_t = \mu + \phi y_{t-1} + u_t, u_t \sim N(0, \sigma_t^2) \quad (8.76)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (8.77)$$

The model is estimated imposing the restriction and the maximised *LLF* falls to 64.54. Is the restriction supported by the data, which would correspond to the situation where an ARCH(1) specification was sufficient? The test statistic is given by

$$LR = -2(64.54 - 66.85) = 4.62 \quad (8.78)$$

The test follows a  $\chi^2(1) = 3.84$  at 5%, so that the null is marginally rejected. It would thus be concluded that an ARCH(1) model, with no lag of the conditional variance in the variance equation, is not quite sufficient to describe the dependence in volatility over time.

### 8.19 Volatility forecasting: some examples and results from the literature

There is a vast and relatively new literature that attempts to compare the accuracies of various models for producing out-of-sample volatility forecasts. Akgiray (1989), for example, finds the GARCH model superior to ARCH, exponentially weighted moving average and historical mean models for forecasting monthly US stock index volatility. A similar result concerning the apparent superiority of GARCH is observed by West and Cho (1995) using one-step-ahead forecasts of dollar exchange rate volatility, although for longer horizons, the model behaves no better than their alternatives. Pagan and Schwert (1990) compare GARCH, EGARCH, Markov switching regime and three non-parametric models for forecasting monthly US stock return volatilities. The EGARCH followed by the GARCH models perform moderately; the remaining models produce very poor predictions. Franses and van Dijk (1996) compare three members of the GARCH family (standard GARCH, QGARCH and the GJR model) for forecasting the weekly volatility of various European stock market indices. They find that the non-linear GARCH models were unable to beat the standard GARCH model. Finally, Brailsford and Faff (1996) find GJR and GARCH models slightly superior to various simpler models for predicting Australian monthly stock index volatility. The conclusion arising from this growing body of research is that forecasting volatility is a 'notoriously difficult task' (Brailsford and Faff, 1996, p. 419), although it appears that conditional heteroscedasticity models are among the best that are currently available. In particular, more complex non-linear and non-parametric models are inferior in prediction to simpler models, a result echoed in an earlier paper by Dimson and Marsh (1990) in the context of relatively complex versus parsimonious linear models. Finally, Brooks (1998), considers whether measures of market volume can assist in improving volatility forecast accuracy, finding that they cannot.

A particularly clear example of the style and content of this class of research is given by Day and Lewis (1992). The Day and Lewis study will therefore now be examined in depth. The purpose of their paper is to consider the out-of-sample forecasting performance of GARCH and EGARCH models for predicting stock index volatility. The forecasts from these econometric models are compared with those given from an 'implied volatility'. As discussed above, implied volatility is the market's expectation of the 'average' level of volatility of an underlying asset over the life of the option that is implied by the current traded price of the option. Given an assumed model for pricing options, such as the Black-Scholes, all of the

inputs to the model except for volatility can be observed directly from the market or are specified in the terms of the option contract. Thus, it is possible, using an iterative search procedure such as the Newton–Raphson method (see, for example, Watsham and Parramore, 2004), to ‘back out’ the volatility of the underlying asset from the option’s price. An important question for research is whether implied or econometric models produce more accurate forecasts of the volatility of the underlying asset. If the options and underlying asset markets are informationally efficient, econometric volatility forecasting models based on past realised values of underlying volatility should have no incremental explanatory power for future values of volatility of the underlying asset. On the other hand, if econometric models do hold additional information useful for forecasting future volatility, it is possible that such forecasts could be turned into a profitable trading rule.

The data employed by Day and Lewis comprise weekly closing prices (Wednesday to Wednesday, and Friday to Friday) for the S&P100 Index option and the underlying index from 11 March 1983–31 December 1989. They employ both mid-week to mid-week returns and Friday to Friday returns to determine whether weekend effects have any significant impact on the latter. They argue that Friday returns contain expiration effects since implied volatilities are seen to jump on the Friday of the week of expiration. This issue is not of direct interest to this book, and consequently only the mid-week to mid-week results will be shown here.

The models that Day and Lewis employ are as follows. First, for the conditional mean of the time series models, they employ a GARCH-M specification for the excess of the market return over a risk-free proxy

$$R_{Mt} - R_{Ft} = \lambda_0 + \lambda_1 \sqrt{h_t} + u_t \quad (8.79)$$

where  $R_{Mt}$  denotes the return on the market portfolio, and  $R_{Ft}$  denotes the risk-free rate. Note that Day and Lewis denote the conditional variance by  $h_t^2$ , while this is modified to the standard  $h_t$  here. Also, the notation  $\sigma_t^2$  will be used to denote implied volatility estimates. For the variance, two specifications are employed: a ‘plain vanilla’ GARCH(1,1) and an EGARCH

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \quad (8.80)$$

or

$$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left( \theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[ \left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| - \left( \frac{2}{\pi} \right)^{1/2} \right] \right) \quad (8.81)$$

One way to test whether implied or GARCH-type volatility models perform best is to add a lagged value of the implied volatility estimate ( $\sigma_{t-1}^2$ ) to (8.80) and (8.81). A ‘hybrid’ or ‘encompassing’ specification would thus result. Equation (8.80) becomes

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \delta \sigma_{t-1}^2 \quad (8.82)$$

and (8.81) becomes

$$\begin{aligned} \ln(h_t) = & \alpha_0 + \beta_1 \ln(h_{t-1}) \\ & + \alpha_1 \left( \theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[ \left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| - \left( \frac{2}{\pi} \right)^{1/2} \right] \right) + \delta \ln(\sigma_{t-1}^2) \end{aligned} \quad (8.83)$$

The tests of interest are given by  $H_0 : \delta = 0$  in (8.82) or (8.83). If these null hypotheses cannot be rejected, the conclusion would be that implied volatility contains no incremental information useful for explaining volatility than that derived from a GARCH model. At the same time,  $H_0 : \alpha_1 = 0$  and  $\beta_1 = 0$  in (8.82), and  $H_0 : \alpha_1 = 0$  and  $\beta_1 = 0$  and  $\theta = 0$  and  $\gamma = 0$  in (8.83) are also tested. If this second set of restrictions holds, then (8.82) and (8.83) collapse to

$$h_t = \alpha_0 + \delta \sigma_{t-1}^2 \quad (8.82')$$

and

$$\ln(h_t) = \alpha_0 + \delta \ln(\sigma_{t-1}^2) \quad (8.83')$$

These sets of restrictions on (8.82) and (8.83) test whether the lagged squared error and lagged conditional variance from a GARCH model contain any additional explanatory power once implied volatility is included in the specification. All of these restrictions can be tested fairly easily using a likelihood ratio test. The results of such a test are presented in table 8.1.

It appears from the coefficient estimates and their standard errors under the specification (8.82) that the implied volatility term ( $\delta$ ) is statistically significant, while the GARCH terms ( $\alpha_1$  and  $\beta_1$ ) are not. However, the test statistics given in the final column are both greater than their corresponding  $\chi^2$  critical values, indicating that both GARCH and implied volatility have incremental power for modelling the underlying stock volatility. A similar analysis is undertaken in Day and Lewis that compares EGARCH with implied volatility. The results are presented here in table 8.2.

The EGARCH results tell a very similar story to those of the GARCH specifications. Neither the lagged information from the EGARCH specification nor the lagged implied volatility terms can be suppressed, according to the

**Table 8.1** GARCH versus implied volatility

$R_{Mt} - R_{Ft} = \lambda_0 + \lambda_1 \sqrt{h_t} + u_t$	(8.79)							
$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}$	(8.80)							
$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \delta \sigma_{t-1}^2$	(8.82)							
$h_t = \alpha_0 + \delta \sigma_{t-1}^2$	(8.82')							
Equation for variance	$\lambda_0$	$\lambda_1$	$\alpha_0 \times 10^{-4}$	$\alpha_1$	$\beta_1$	$\delta$	Log-L	$\chi^2$
(8.80)	0.0072 (0.005)	0.071 (0.01)	5.428 (1.65)	0.093 (0.84)	0.854 (8.17)	—	767.321	17.77
(8.82)	0.0015 (0.028)	0.043 (0.02)	2.065 (2.98)	0.266 (1.17)	-0.068 (-0.59)	0.318 (3.00)	776.204	—
(8.82')	0.0056 (0.001)	-0.184 (-0.001)	0.993 (1.50)	—	—	0.581 (2.94)	764.394	23.62

Notes: *t*-ratios in parentheses, Log-L denotes the maximised value of the log-likelihood function in each case.  $\chi^2$  denotes the value of the test statistic, which follows a  $\chi^2(1)$  in the case of (8.82) restricted to (8.80), and a  $\chi^2(2)$  in the case of (8.82) restricted to (8.82').

Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

**Table 8.2** EGARCH versus implied volatility

$R_{Mt} - R_{Ft} = \lambda_0 + \lambda_1 \sqrt{h_t} + u_t$	(8.79)								
$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left( \theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[ \left  \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right  - \left( \frac{2}{\pi} \right)^{1/2} \right] \right)$	(8.81)								
$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left( \theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[ \left  \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right  - \left( \frac{2}{\pi} \right)^{1/2} \right] \right) + \delta \ln(\sigma_{t-1}^2)$	(8.83)								
$\ln(h_t) = \alpha_0 + \delta \ln(\sigma_{t-1}^2)$	(8.83')								
Equation for variance	$\lambda_0$	$\lambda_1$	$\alpha_0 \times 10^{-4}$	$\beta_1$	$\theta$	$\gamma$	$\delta$	Log-L	$\chi^2$
(8.81)	-0.0026 (-0.03)	0.094 (0.25)	-3.62 (-2.90)	0.529 (3.26)	0.273 (-4.13)	0.357 (3.17)	—	776.436	8.09
(8.83)	0.0035 (0.56)	-0.076 (-0.24)	-2.28 (-1.82)	0.373 (1.48)	-0.282 (-4.34)	0.210 (1.89)	0.351 (1.82)	780.480	—
(8.83')	0.0047 (0.71)	-0.139 (-0.43)	-2.76 (-2.30)	—	—	—	0.667 (4.01)	765.034	30.89

Notes: *t*-ratios in parentheses, Log-L denotes the maximised value of the log-likelihood function in each case.  $\chi^2$  denotes the value of the test statistic, which follows a  $\chi^2(1)$  in the case of (8.83) restricted to (8.81), and a  $\chi^2(3)$  in the case of (8.83) restricted to (8.83').

Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

likelihood ratio statistics. In specification (8.83), both the EGARCH terms and the implied volatility coefficients are marginally significant.

However, the tests given above do not represent a true test of the predictive ability of the models, since all of the observations were used in both estimating and testing the models. Hence the authors proceed to conduct an out-of-sample forecasting test. There are a total of 729 data points in their sample. They use the first 410 to estimate the models, and then make a one-step-ahead forecast of the following week's volatility. They then roll the sample forward one observation at a time, constructing a new one-step-ahead forecast at each stage.

They evaluate the forecasts in two ways. The first is by regressing the realised volatility series on the forecasts plus a constant

$$\sigma_{t+1}^2 = b_0 + b_1\sigma_{f_t}^2 + \xi_{t+1} \quad (8.84)$$

where  $\sigma_{t+1}^2$  is the 'actual' value of volatility at time  $t + 1$ , and  $\sigma_{f_t}^2$  is the value forecasted for it during period  $t$ . Perfectly accurate forecasts would imply  $b_0 = 0$  and  $b_1 = 1$ . The second method is via a set of forecast encompassing tests. Essentially, these operate by regressing the realised volatility on the forecasts generated by several models. The forecast series that have significant coefficients are concluded to encompass those of models whose coefficients are not significant.

But what is volatility? In other words, with what measure of realised or 'ex post' volatility should the forecasts be compared? This is a question that received very little attention in the literature until recently. A common method employed is to assume, for a daily volatility forecasting exercise, that the relevant *ex post* measure is the square of that day's return. For any random variable  $r_t$ , its conditional variance can be expressed as

$$\text{var}(r_t) = E[r_t - E(r_t)]^2 \quad (8.85)$$

As stated previously, it is typical, and not unreasonable for relatively high frequency data, to assume that  $E(r_t)$  is zero, so that the expression for the variance reduces to

$$\text{var}(r_t) = E[r_t^2] \quad (8.86)$$

Andersen and Bollerslev (1998) argue that squared daily returns provide a very noisy proxy for the true volatility, and a much better proxy for the day's variance would be to compute the volatility for the day from intra-daily data. For example, a superior daily variance measure could be obtained by taking hourly returns, squaring them and adding them up. The reason that the use of higher frequency data provides a better measure of *ex post* volatility is simply that it employs more information.

By using only daily data to compute a daily volatility measure, effectively only two observations on the underlying price series are employed. If the daily closing price is the same one day as the next, the squared return and therefore the volatility would be calculated to be zero, when there may have been substantial intra-day fluctuations. Hansen and Lunde (2006) go further and suggest that even the ranking of models by volatility forecast accuracy could be inconsistent if the evaluation uses a poor proxy for the true, underlying volatility.

Day and Lewis use two measures of *ex post* volatility in their study (for which the frequency of data employed in the models is weekly):

- (1) The square of the weekly return on the index, which they call SR
- (2) The variance of the week's daily returns multiplied by the number of trading days in that week, which they call WV.

The Andersen and Bollerslev argument implies that the latter measure is likely to be superior, and therefore that more emphasis should be placed on those results.

The results for the separate regressions of realised volatility on a constant and the forecast are given in table 8.3.

The coefficient estimates for  $b_0$  given in table 8.3 can be interpreted as indicators of whether the respective forecasting approaches are biased. In all cases, the  $b_0$  coefficients are close to zero. Only for the historic volatility forecasts and the implied volatility forecast when the *ex post* measure is the squared weekly return, are the estimates statistically significant. Positive coefficient estimates would suggest that on average the forecasts are too low. The estimated  $b_1$  coefficients are in all cases a long way from unity, except for the GARCH (with daily variance *ex post* volatility) and EGARCH (with squared weekly variance as *ex post* measure) models. Finally, the  $R^2$  values are very small (all less than 10%, and most less than 3%), suggesting that the forecast series do a poor job of explaining the variability of the realised volatility measure.

The forecast encompassing regressions are based on a procedure due to Fair and Shiller (1990) that seeks to determine whether differing sets of forecasts contain different sets of information from one another. The test regression is of the form

$$\sigma_{t+1}^2 = b_0 + b_1\sigma_{I_t}^2 + b_2\sigma_{G_t}^2 + b_3\sigma_{E_t}^2 + b_4\sigma_{H_t}^2 + \xi_{t+1} \quad (8.87)$$

with results presented in table 8.4.

The sizes and significances of the coefficients in table 8.4 are of interest. The most salient feature is the lack of significance of most of the forecast series. In the first comparison, neither the implied nor the GARCH

**Table 8.3** Out-of-sample predictive power for weekly volatility forecasts

$\sigma_{t+1}^2 = b_0 + b_1\sigma_{f_t}^2 + \xi_{t+1}$ (8.84)				
Forecasting model	Proxy for <i>ex post</i> volatility	$b_0$	$b_1$	$R^2$
Historic	SR	0.0004 (5.60)	0.129 (21.18)	0.094
Historic	WV	0.0005 (2.90)	0.154 (7.58)	0.024
GARCH	SR	0.0002 (1.02)	0.671 (2.10)	0.039
GARCH	WV	0.0002 (1.07)	1.074 (3.34)	0.018
EGARCH	SR	0.0000 (0.05)	1.075 (2.06)	0.022
EGARCH	WV	-0.0001 (-0.48)	1.529 (2.58)	0.008
Implied volatility	SR	0.0022 (2.22)	0.357 (1.82)	0.037
Implied volatility	WV	0.0005 (0.389)	0.718 (1.95)	0.026

Notes: 'Historic' refers to the use of a simple historical average of the squared returns to forecast volatility; *t*-ratios in parentheses; SR and WV refer to the square of the weekly return on the S&P100, and the variance of the week's daily returns multiplied by the number of trading days in that week, respectively.

Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

**Table 8.4** Comparisons of the relative information content of out-of-sample volatility forecasts

$\sigma_{t+1}^2 = b_0 + b_1\sigma_{I_t}^2 + b_2\sigma_{G_t}^2 + b_3\sigma_{E_t}^2 + b_4\sigma_{H_t}^2 + \xi_{t+1}$ (8.87)						
Forecast comparisons	$b_0$	$b_1$	$b_2$	$b_3$	$b_4$	$R^2$
Implied versus GARCH	-0.00010 (-0.09)	0.601 (1.03)	0.298 (0.42)	-	-	0.027
Implied versus GARCH versus Historical	0.00018 (1.15)	0.632 (1.02)	-0.243 (-0.28)	-	0.123 (7.01)	0.038
Implied versus EGARCH	-0.00001 (-0.07)	0.695 (1.62)	-	0.176 (0.27)	-	0.026
Implied versus EGARCH versus Historical	0.00026 (1.37)	0.590 (1.45)	-0.374 (-0.57)	-	0.118 (7.74)	0.038
GARCH versus EGARCH	0.00005 (0.370)	-	1.070 (2.78)	-0.001 (-0.00)	-	0.018

Notes: *t*-ratios in parentheses; the *ex post* measure used in this table is the variance of the week's daily returns multiplied by the number of trading days in that week.

Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

forecast series have statistically significant coefficients. When historical volatility is added, its coefficient is positive and statistically significant. An identical pattern emerges when forecasts from implied and EGARCH models are compared: that is, neither forecast series is significant, but when a simple historical average series is added, its coefficient is significant. It is clear from this, and from the last row of table 8.4, that the asymmetry term in the EGARCH model has no additional explanatory power compared with that embodied in the symmetric GARCH model. Again, all of the  $R^2$  values are very low (less than 4%).

The conclusion reached from this study (which is broadly in line with many others) is that within sample, the results suggest that implied volatility contains extra information not contained in the GARCH/EGARCH specifications. But the out-of-sample results suggest that predicting volatility is a difficult task!

## 8.20 Stochastic volatility models revisited

Under the heading of models for time-varying volatilities, only approaches based on the GARCH class of models have been discussed thus far. Another class of models is also available, known as stochastic volatility (SV) models. It is a common misconception that GARCH-type specifications are sorts of stochastic volatility models. However, as the name suggests, stochastic volatility models differ from GARCH principally in that the conditional variance equation of a GARCH specification is completely deterministic given all information available up to that of the previous period. In other words, there is no error term in the variance equation of a GARCH model, only in the mean equation.

Stochastic volatility models contain a second error term, which enters into the conditional variance equation. A very simple example of a stochastic volatility model would be the autoregressive volatility specification described in section 8.6. This model is simple to understand and simple to estimate, because it requires that we have an observable measure of volatility which is then simply used as any other variable in an autoregressive model. However, the term ‘stochastic volatility’ is usually associated with a different formulation, a possible example of which would be

$$y_t = \mu + u_t \sigma_t, u_t \sim N(0, 1) \quad (8.88)$$

$$\log(\sigma_t^2) = \alpha_0 + \beta_1 \log(\sigma_{t-1}^2) + \sigma_\eta \eta_t \quad (8.89)$$

where  $\eta_t$  is another  $N(0,1)$  random variable that is independent of  $u_t$ . Here the volatility is latent rather than observed, and so is modelled indirectly.

Stochastic volatility models are closely related to the financial theories used in the options pricing literature. Early work by Black and Scholes (1973) had assumed that volatility is constant through time. Such an assumption was made largely for simplicity, although it could hardly be considered realistic. One unappealing side-effect of employing a model with the embedded assumption that volatility is fixed, is that options deep in-the-money and far out-of-the-money are underpriced relative to actual traded prices. This empirical observation provided part of the genesis for stochastic volatility models, where the logarithm of an unobserved variance process is modelled by a linear stochastic specification, such as an autoregressive model. The primary advantage of stochastic volatility models is that they can be viewed as discrete time approximations to the continuous time models employed in options pricing frameworks (see, for example, Hull and White, 1987). However, such models are hard to estimate. For reviews of (univariate) stochastic volatility models, see Taylor (1994), Ghysels *et al.* (1995) or Shephard (1996) and the references therein.

While stochastic volatility models have been widely employed in the mathematical options pricing literature, they have not been popular in empirical discrete-time financial applications, probably owing to the complexity involved in the process of estimating the model parameters (see Harvey, Ruiz and Shephard, 1994). So, while GARCH-type models are further from their continuous time theoretical underpinnings than stochastic volatility, they are much simpler to estimate using maximum likelihood. A relatively simple modification to the maximum likelihood procedure used for GARCH model estimation is not available, and hence stochastic volatility models are not discussed further here.

## 8.21 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. 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. 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 incorporating 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.

## 8.22 Covariance modelling and forecasting in finance: some examples

### 8.22.1 *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

$$\beta_{i,t} = \frac{\sigma_{im,t}}{\sigma_{m,t}^2} \quad (8.90)$$

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  $t$ .

## 8.22.2 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 (2005) by first defining:

$\Delta S$  = change in spot price  $S$ , during the life of the hedge  
 $\Delta F$  = change in futures price,  $F$ , during the life of the hedge  
 $\sigma_s$  = standard deviation of  $\Delta S$   
 $\sigma_F$  = standard deviation of  $\Delta F$   
 $\rho$  = correlation coefficient between  $\Delta S$  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

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

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

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

or

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

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

$$v = \sigma_s^2 + h^2\sigma_F^2 - 2hp\sigma_s\sigma_F \quad (8.91)$$

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

$$h = \rho \frac{\sigma_s}{\sigma_F} \quad (8.92)$$

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

$$h_t = p_t \frac{\sigma_{s,t}}{\sigma_{F,t}} \quad (8.93)$$

Various models are available for covariance or correlation forecasting, and several will be discussed below.

### 8.23 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.

### 8.24 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

$$\tilde{\sigma}^2(xy) = \tilde{\sigma}^2(x) + \tilde{\sigma}^2(y) - 2\tilde{\sigma}(x, y) \quad (8.94)$$

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 (8.94), the implied covariance is obtained via

$$\tilde{\sigma}(x, y) = \frac{\tilde{\sigma}^2(x) + \tilde{\sigma}^2(y) - \tilde{\sigma}^2(xy)}{2} \quad (8.95)$$

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 (8.94).

## 8.25 Exponentially weighted moving average model for covariances

Again, as for the case of volatility modelling, an 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 estimate for covariance at time  $t$  and the forecast for subsequent periods may be written

$$\sigma(x, y)_t = (1 - \lambda) \sum_{i=0}^{\infty} \lambda^i x_{t-i} y_{t-i} \quad (8.96)$$

with  $\lambda(0 < \lambda < 1)$  again denoting the decay factor, determining the relative weights attached to recent versus less recent observations.

## 8.26 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. Several different multivariate GARCH formulations have been proposed in the literature, including the *VECH*, the diagonal *VECH* and the *BEKK* models. Each of these is discussed in turn below; for a more detailed discussion, see Kroner and Ng (1998). In each case, it is assumed below for simplicity that there are two assets, whose return variances and covariances are to be modelled. For an excellent survey of multivariate GARCH models, see Bauwens, Laurent and Rombouts (2006).<sup>2</sup>

### 8.26.1 The *VECH* model

A common specification of the *VECH* model, initially due to Bollerslev, Engle and Wooldridge (1988), is

$$VECH(H_t) = C + AVECH(\Xi_{t-1} \Xi'_{t-1}) + BVECH(H_{t-1}) \\ \Xi_t | \psi_{t-1} \sim N(0, H_t), \quad (8.97)$$

where  $H_t$  is a  $2 \times 2$  conditional variance–covariance matrix,  $\Xi_t$  is a  $2 \times 1$  innovation (disturbance) vector,  $\psi_{t-1}$  represents the information set at time  $t - 1$ ,  $C$  is a  $3 \times 1$  parameter vector,  $A$  and  $B$  are  $3 \times 3$  parameter matrices and  $VECH(\cdot)$  denotes the column-stacking operator applied to the upper portion of the symmetric matrix. The model requires the estimation

<sup>2</sup> It is also worth noting that there also exists a class of multivariate stochastic volatility models. These were originally proposed by Harvey, Ruiz and Shephard (1994), although see also Brooks (2006).

of 21 parameters ( $C$  has 3 elements,  $A$  and  $B$  each have 9 elements). In order to gain a better understanding of how the  $VECH$  model works, the elements 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  $h_{ii}$  represent the conditional variances at time  $t$  of the two-asset return series ( $i = 1, 2$ ) used in the model, and  $h_{ij}$  ( $i \neq j$ ) represent the conditional covariances between the asset returns. In the case of  $VECH(\Xi_t \Xi_t')$ , this can be expressed as

$$\begin{aligned} VECH(\Xi_t \Xi_t') &= VECH\left(\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \begin{bmatrix} u_{1t} & u_{2t} \end{bmatrix}\right) \\ &= VECH\left(\begin{array}{cc} u_{1t}^2 & u_{1t}u_{2t} \\ u_{1t}u_{2t} & u_{2t}^2 \end{array}\right) \\ &= \begin{bmatrix} u_{1t}^2 \\ u_{2t}^2 \\ u_{1t}u_{2t} \end{bmatrix} \end{aligned}$$

The  $VECH$  model in full is given by

$$\begin{aligned} h_{11t} &= c_{11} + a_{11}u_{1t-1}^2 + a_{12}u_{2t-1}^2 + a_{13}u_{1t-1}u_{2t-1} + b_{11}h_{11t-1} \\ &\quad + b_{12}h_{22t-1} + b_{13}h_{12t-1} \end{aligned} \quad (8.98)$$

$$\begin{aligned} h_{22t} &= c_{21} + a_{21}u_{1t-1}^2 + a_{22}u_{2t-1}^2 + a_{23}u_{1t-1}u_{2t-1} + b_{21}h_{11t-1} \\ &\quad + b_{22}h_{22t-1} + b_{23}h_{12t-1} \end{aligned} \quad (8.99)$$

$$\begin{aligned} h_{12t} &= c_{31} + a_{31}u_{1t-1}^2 + a_{32}u_{2t-1}^2 + a_{33}u_{1t-1}u_{2t-1} + b_{31}h_{11t-1} \\ &\quad + b_{32}h_{22t-1} + b_{33}h_{12t-1} \end{aligned} \quad (8.100)$$

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. Estimation of such a model would be quite a formidable task, even in the two-asset case considered here.

### 8.26.2 The diagonal VECH model

Even in the simple case of two assets, the conditional variance and covariance equations for the unrestricted VECH model contain 21 parameters. As the number of assets employed in the model increases, the 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 reduces the number of parameters to be estimated to 9 (now  $A$  and  $B$  each have 3 elements) and the model, known as a diagonal VECH, is now characterised by

$$h_{ij,t} = \omega_{ij} + \alpha_{ij}u_{i,t-1}u_{j,t-1} + \beta_{ij}h_{ij,t-1} \quad \text{for } i, j = 1, 2, \quad (8.101)$$

where  $\omega_{ij}$ ,  $\alpha_{ij}$  and  $\beta_{ij}$  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 or factor GARCH models (see Alexander, 2001). 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.

### 8.26.3 The BEKK model

The BEKK model (Engle and Kroner, 1995) addresses the difficulty with VECM of ensuring that the  $H$  matrix is always positive definite. It is represented by

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi'_{t-1}B \quad (8.102)$$

where  $A$ , and  $B$  are  $2 \times 2$  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.

### 8.26.4 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

$$\ell(\theta) = -\frac{TN}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T (\log |H_t| + \Xi_t' H_t^{-1} \Xi_t) \quad (8.103)$$

where  $\theta$  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  $\theta$  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.

## 8.27 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 6-month US Treasury bills, 20-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

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} 0.070 \\ (0.032) \\ -4.342 \\ (1.030) \\ -3.117 \\ (0.710) \end{pmatrix} + 0.499 \sum_j \omega_{jt-1} \begin{pmatrix} h_{1jt} \\ h_{2jt} \\ h_{3jt} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} \quad (8.104)$$

$$\begin{pmatrix} h_{11t} \\ h_{12t} \\ h_{22t} \\ h_{13t} \\ h_{23t} \\ h_{33t} \end{pmatrix} = \begin{pmatrix} 0.011 \\ (0.004) \\ 0.176 \\ (0.062) \\ 13.305 \\ (6.372) \\ 0.018 \\ (0.009) \\ 5.143 \\ (2.820) \\ 2.083 \\ (1.466) \end{pmatrix} + \begin{pmatrix} 0.445\varepsilon_{1t-1}^2 \\ (0.105) \\ 0.233\varepsilon_{1t-1}\varepsilon_{2t-1} \\ (0.092) \\ 0.188\varepsilon_{2t-1}^2 \\ (0.113) \\ 0.197\varepsilon_{1t-1}\varepsilon_{3t-1} \\ (0.132) \\ 0.165\varepsilon_{2t-1}\varepsilon_{3t-1} \\ (0.093) \\ 0.078\varepsilon_{3t-1}^2 \\ (0.066) \end{pmatrix} + \begin{pmatrix} 0.466h_{11t-1} \\ (0.056) \\ 0.598h_{12t-1} \\ (0.052) \\ 0.441h_{22t-1} \\ (0.215) \\ -0.362h_{13t-1} \\ (0.361) \\ -0.348h_{23t-1} \\ (0.338) \\ 0.469h_{33t-1} \\ (0.333) \end{pmatrix} \quad (8.105)$$

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

where  $y_{jt}$  are the returns,  $\omega_{jt-1}$  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 (8.104) and (8.105). 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.

## **8.28 Estimating a time-varying hedge ratio for FTSE stock index returns**

A paper by Brooks, Henry and Persaud (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.

### *8.28.1 Background*

There has been much empirical research into the calculation of optimal hedge ratios. The general consensus is that the use of multivariate generalised autoregressive conditionally heteroscedastic (MGARCH) models yields superior performances, evidenced by lower portfolio volatilities, than either time-invariant or rolling ordinary least squares (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.

### 8.28.2 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  $\beta$  units of the futures contract may be written as

$$E_{t-1}(R_t) = E_{t-1}(\Delta S_t) - \beta_{t-1} E_{t-1}(\Delta F_t) \quad (8.106)$$

where  $\beta_{t-1}$  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

$$h_{p,t} = h_{s,t} + \beta_{t-1}^2 h_{F,t} - 2\beta_{t-1} h_{SF,t} \quad (8.107)$$

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.  $\beta_{t-1}^*$ , the optimal number of futures contracts in the investor's portfolio, i.e. the optimal hedge ratio, is given by

$$\beta_{t-1}^* = -\frac{h_{SF,t}}{h_{F,t}} \quad (8.108)$$

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

$$\Delta S_t = a + b\Delta F_t + u_t \quad (8.109)$$

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

### 8.28.3 Data and results

The data employed in the Brooks, Henry and Persaud (2002) study comprises 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 1-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

**Table 8.5** Hedging effectiveness: summary statistics for portfolio returns

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

Note: *t*-ratios displayed as {.}.

Source: Brooks, Henry and Persaud (2002).

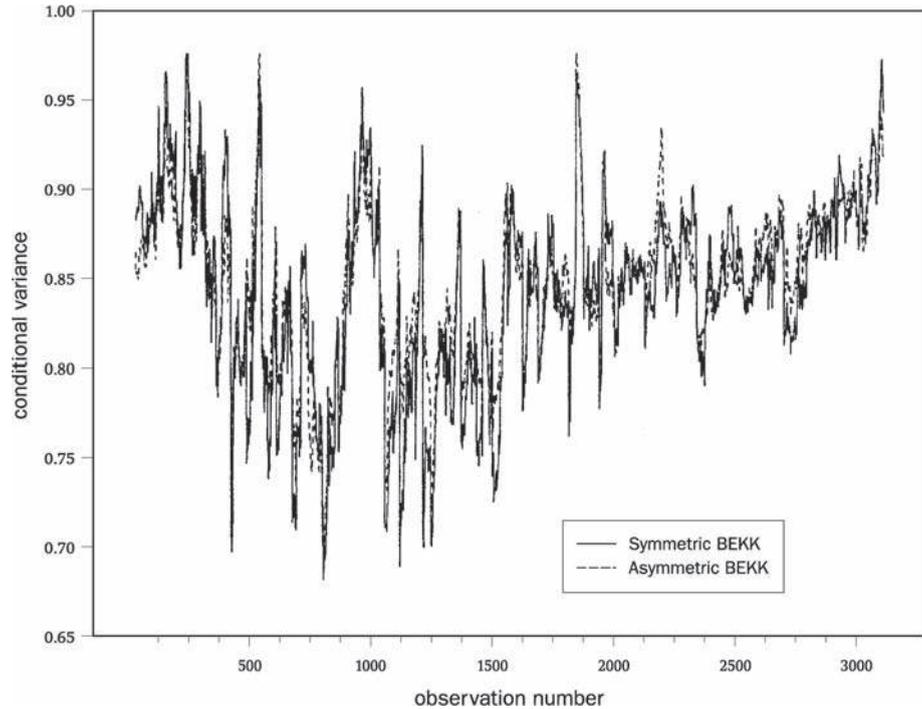
ratios using (8.108). The hedging performance of a BEKK formulation is examined, and also a BEKK model including asymmetry terms (in the same style as GJR models). The returns and variances for the various hedging strategies are presented in table 8.5.

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 8.5 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

**Figure 8.5**

Source: Brooks, Henry and Persaud (2002). Time-varying hedge ratios derived from symmetric and asymmetric BEKK models for FTSE returns.



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 8.5 graphs the time-varying hedge ratio from the symmetric and asymmetric MGARCH models. 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  $\beta_{t-1}^* \sim I(1)$  (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.

Brooks, Henry and Persaud 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.

## 8.29 Estimating multivariate GARCH models using EViews

In previous versions of the software, multivariate GARCH models could only be estimated in EViews by writing the required instructions, but now they are available using the menus. To estimate such a model, first you need to create a system that contains the variables to be used. **Highlight the three variables 'reur', 'rgbp', and 'rjpy'** and then **right click the mouse**. Choose **Open/as System...**; **Click Object/New Object** and then click **System**. Screenshot 8.6 will appear.

**Screenshot 8.6**

Making a system

The screenshot shows the 'Make System' dialog box in EViews. The 'Dependent variables' field contains the list 'reur rgbp rjpy'. The 'Regressors and AR() terms' section has 'Common coefficients' empty and 'Equation specific coefficients' containing 'c'. The 'Instrument list' section has 'Common' empty and 'Equation specific' empty. The 'Coefficient name' field contains 'C'. The 'Option' dropdown menu is set to 'NONE'. The 'OK' and 'Cancel' buttons are visible at the bottom right.

Since no explanatory variables will be used in the conditional mean equation, all of the default choices can be retained, so just click OK. A system box containing the three equations with just intercepts will be seen. Then click **Proc/Estimate...** for the 'System Estimation' window. Change the 'Estimation method' to **ARCH - Conditional Heteroscedasticity**. EViews permits the estimation of 3 important classes of multivariate GARCH model: the diagonal VEC, the constant conditional correlation,

and the diagonal BEKK models. For the error distribution, either a multivariate normal or a multivariate Student's  $t$  can be used. Additional exogenous variables can be incorporated into the variance equation, and asymmetries can be allowed for. Leaving all of these options as the defaults and clicking OK would yield the following results.<sup>3</sup>

System: UNTITLED  
 Estimation Method: ARCH Maximum Likelihood (Marquardt)  
 Covariance specification: Diagonal VECH  
 Date: 09/06/07 Time: 20:27  
 Sample: 7/08/2002 7/07/2007  
 Included observations: 1826  
 Total system (balanced) observations 5478  
 Presample covariance: backcast (parameter = 0.7)  
 Convergence achieved after 97 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	-0.024107	0.008980	-2.684689	0.0073
C(2)	-0.014243	0.008861	-1.607411	0.1080
C(3)	0.005420	0.009368	0.578572	0.5629
Variance Equation Coefficients				
C(4)	0.006725	0.000697	9.651785	0.0000
C(5)	0.054984	0.004840	11.36043	0.0000
C(6)	0.004792	0.000979	4.895613	0.0000
C(7)	0.129606	0.007495	17.29127	0.0000
C(8)	0.030076	0.003945	7.624554	0.0000
C(9)	0.006344	0.001276	4.971912	0.0000
C(10)	0.031130	0.002706	11.50347	0.0000
C(11)	0.047425	0.004734	10.01774	0.0000
C(12)	0.022325	0.004061	5.497348	0.0000
C(13)	0.121511	0.012267	9.905618	0.0000
C(14)	0.059994	0.007375	8.135074	0.0000
C(15)	0.034482	0.005079	6.788698	0.0000
C(16)	0.937158	0.004929	190.1436	0.0000
C(17)	0.560650	0.034187	16.39950	0.0000
C(18)	0.933618	0.011479	81.33616	0.0000
C(19)	0.127121	0.039195	3.243308	0.0012
C(20)	0.582251	0.047292	12.31189	0.0000
C(21)	0.931788	0.010298	90.47833	0.0000
Log likelihood	-1935.756	Schwarz criterion		2.206582
Avg. log likelihood	-0.353369	Hannan-Quinn criter.		2.166590
Akaike info criterion	2.143216			

<sup>3</sup> The complexity of this model means that it takes longer to estimate than any of the univariate GARCH or other models examined previously.

Transformed Variance Coefficients				
	Coefficient	Std. Error	z-Statistic	Prob.
M(1,1)	0.006725	0.000697	9.651785	0.0000
M(1,2)	0.054984	0.004840	11.36043	0.0000
M(1,3)	0.004792	0.000979	4.895613	0.0000
M(2,2)	0.129606	0.007495	17.29127	0.0000
M(2,3)	0.030076	0.003945	7.624554	0.0000
M(3,3)	0.006344	0.001276	4.971912	0.0000
A1(1,1)	0.031130	0.002706	11.50347	0.0000
A1(1,2)	0.047425	0.004734	10.01774	0.0000
A1(1,3)	0.022325	0.004061	5.497348	0.0000
A1(2,2)	0.121511	0.012267	9.905618	0.0000
A1(2,3)	0.059994	0.007375	8.135074	0.0000
A1(3,3)	0.034482	0.005079	6.788698	0.0000
B1(1,1)	0.937158	0.004929	190.1436	0.0000
B1(1,2)	0.560650	0.034187	16.39950	0.0000
B1(1,3)	0.933618	0.011479	81.33616	0.0000
B1(2,2)	0.127121	0.039195	3.243308	0.0012
B1(2,3)	0.582251	0.047292	12.31189	0.0000
B1(3,3)	0.931788	0.010298	90.47833	0.0000

Equation: REUR = C(1)			
R-squared	-0.000151	Mean dependent var	-0.018327
Adjusted R-squared	-0.000151	S.D. dependent var	0.469930
S.E. of regression	0.469965	Sum squared resid	403.0827
Prob(F-statistic)	2.050379		

Equation: RGBP = C(2)			
R-squared	-0.000006	Mean dependent var	-0.015282
Adjusted R-squared	-0.000006	S.D. dependent var	0.413105
S.E. of regression	0.413106	Sum squared resid	311.4487
Prob(F-statistic)	1.918603		

Equation: RJPY = C(3)			
R-squared	-0.000087	Mean dependent var	0.001328
Adjusted R-squared	-0.000087	S.D. dependent var	0.439632
S.E. of regression	0.439651	Sum squared resid	352.7596
Prob(F-statistic)	1.981767		

Covariance specification: Diagonal VECM  
 GARCH = M + A1.\*RESID(-1)\*RESID(-1)' + B1.\*GARCH(-1)  
 M is an indefinite matrix  
 A1 is an indefinite matrix  
 B1 is an indefinite matrix

The first panel of the table presents the conditional mean estimates; in this example, only intercepts were used in the mean equations. The next panel shows the variance equation coefficients, followed by some measures of goodness of fit for the model as a whole and then for each individual mean equation. The final panel presents the transformed variance coefficients, which in this case are identical to the panel of variance coefficients since no transformation is conducted with normal errors (these would only be different if a Student's  $t$  specification were used). It is evident that the parameter estimates are all both plausible and statistically significant.

There are a number of useful further steps that can be conducted once the model has been estimated, all of which are available by clicking the 'View' button. For example, we can plot the series of residuals, or estimate the correlations between them. Or by clicking on 'Conditional variance', we can list or plot the values of the conditional variances and covariances over time. We can also test for autocorrelation and normality of the errors.

### 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 VEC
- news impact curve
- volatility clustering
- GARCH model
- Wald test
- likelihood ratio test
- GJR specification
- exponentially weighted moving average
- BEKK model
- GARCH-in-mean

## Appendix: 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

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad (8A.1)$$

Assuming that  $u_t \sim N(0, \sigma^2)$ , then  $y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$  so that the probability density function for a normally distributed random variable with

this mean and variance is given by

$$f(y_t | \beta_1 + \beta_2 x_t, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \right\} \quad (8A.2)$$

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  $y$ s are iid, the joint probability density function (pdf) for all the  $y$ s can be expressed as a product of the individual density functions

$$\begin{aligned} & f(y_1, y_2, \dots, y_T | \beta_1 + \beta_2 x_1, \beta_1 + \beta_2 x_2, \dots, \beta_1 + \beta_2 x_T, \sigma^2) \\ &= f(y_1 | \beta_1 + \beta_2 x_1, \sigma^2) f(y_2 | \beta_1 + \beta_2 x_2, \sigma^2) \dots f(y_T | \beta_1 + \beta_2 x_T, \sigma^2) \\ &= \prod_{t=1}^T f(y_t | \beta_1 + \beta_2 x_t, \sigma^2) \quad \text{for } t = 1, \dots, T \end{aligned} \quad (8A.3)$$

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 (8A.3) shows the probability of obtaining all of the values of  $y$  that did occur. Substituting into (8A.3) for every  $y_t$  from (8A.2), and using the result that  $Ae^{x_1} \times Ae^{x_2} \times \dots \times Ae^{x_T} = A^T (e^{x_1} \times e^{x_2} \times \dots \times e^{x_T}) = A^T e^{(x_1+x_2+\dots+x_T)}$ , the following expression is obtained

$$\begin{aligned} & f(y_1, y_2, \dots, y_T | \beta_1 + \beta_2 x_t, \sigma^2) \\ &= \frac{1}{\sigma^T (\sqrt{2\pi})^T} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \right\} \end{aligned} \quad (8A.4)$$

This is the joint density of all of the  $y$ s given the values of  $x_t$ ,  $\beta_1$ ,  $\beta_2$  and  $\sigma^2$ . 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  $\beta_1$ ,  $\beta_2$ ,  $\sigma^2$  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

$$LF(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^T (\sqrt{2\pi})^T} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \right\} \quad (8A.5)$$

Maximum likelihood estimation involves choosing parameter values ( $\beta_1$ ,  $\beta_2$ ,  $\sigma^2$ ) 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 (8A.5) w.r.t.  $\beta_1$ ,  $\beta_2$ ,  $\sigma^2$ , but (8A.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 (8A.5) 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

$$LLF = -T \ln \sigma - \frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \quad (8A.6)$$

which is equivalent to

$$LLF = -\frac{T}{2} \ln \sigma^2 - \frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \quad (8A.7)$$

Only the first part of the RHS of (8A.6) has been changed in (8A.7) to make  $\sigma^2$  appear in that part of the expression rather than  $\sigma$ .

Remembering the result that

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

and differentiating (8A.7) w.r.t.  $\beta_1$ ,  $\beta_2$ ,  $\sigma^2$ , the following expressions for the first derivatives are obtained

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

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

$$\frac{\partial LLF}{\partial \sigma^2} = -\frac{T}{2} \frac{1}{\sigma^2} + \frac{1}{2} \sum \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^4} \quad (8A.10)$$

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

$$\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) = 0 \quad (8A.11)$$

$$\sum y_t - \sum \hat{\beta}_1 - \sum \hat{\beta}_2 x_t = 0 \quad (8A.12)$$

$$\sum y_t - T \hat{\beta}_1 - \hat{\beta}_2 \sum x_t = 0 \quad (8A.13)$$

$$\frac{1}{T} \sum y_t - \hat{\beta}_1 - \hat{\beta}_2 \frac{1}{T} \sum x_t = 0 \quad (8A.14)$$

Recall that

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

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

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \quad (8A.15)$$

From (8A.9)

$$\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) x_t = 0 \quad (8A.16)$$

$$\sum y_t x_t - \sum \hat{\beta}_1 x_t - \sum \hat{\beta}_2 x_t^2 = 0 \quad (8A.17)$$

$$\sum y_t x_t - \hat{\beta}_1 \sum x_t - \hat{\beta}_2 \sum x_t^2 = 0 \quad (8A.18)$$

$$\hat{\beta}_2 \sum x_t^2 = \sum y_t x_t - (\bar{y} - \hat{\beta}_2 \bar{x}) \sum x_t \quad (8A.19)$$

$$\hat{\beta}_2 \sum x_t^2 = \sum y_t x_t - T \bar{x} \bar{y} + \hat{\beta}_2 T \bar{x}^2 \quad (8A.20)$$

$$\hat{\beta}_2 (\sum x_t^2 - T \bar{x}^2) = \sum y_t x_t - T \bar{x} \bar{y} \quad (8A.21)$$

$$\hat{\beta}_2 = \frac{\sum y_t x_t - T \bar{x} \bar{y}}{(\sum x_t^2 - T \bar{x}^2)} \quad (8A.22)$$

From (8A.10)

$$\frac{T}{\hat{\sigma}^2} = \frac{1}{\hat{\sigma}^4} \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2 \quad (8A.23)$$

Rearranging,

$$\hat{\sigma}^2 = \frac{1}{T} \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2 \quad (8A.24)$$

But the term in parentheses on the RHS of (8A.24) is the residual for time  $t$  (i.e. the actual minus the fitted value), so

$$\hat{\sigma}^2 = \frac{1}{T} \sum \hat{u}_t^2 \quad (8A.25)$$

How do these formulae compare with the OLS estimators? (8A.15) and (8A.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 (8A.25) is different. The OLS estimator was

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum \hat{u}_t^2 \quad (8A.26)$$

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.

## Review 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( $\rho$ )?
- (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

$$y_t = \mu + u_t, \quad u_t \sim N(0, \sigma_t^2) \quad (8.110)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (8.111)$$

If  $y_t$  is a daily stock return series, what range of values are likely for the coefficients  $\mu$ ,  $\alpha_0$ ,  $\alpha_1$  and  $\beta$ ?

- (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  $\sigma_t^2$  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 re-considering 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. 1-step-ahead volatility forecasts  
 ii. 20-step-ahead volatility forecasts.  
 (a) If  $u_t$  follows a GARCH(1,1) process, what would be the likely result if a regression of the form (8.110) were estimated using OLS and assuming a constant conditional variance?  
 (b) 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

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \quad (8.112)$$

$$\begin{aligned} \log(\sigma_t^2) = & \alpha_0 + \alpha_1 \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha_2 \log(\sigma_{t-1}^2) \\ & + \alpha_3 \left[ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \end{aligned} \quad (8.113)$$

	GARCH	EGARCH
$\mu$	-0.0130 (0.0669)	-0.0278 (0.0855)
$\alpha_0$	0.0019 (0.0017)	0.0823 (0.5728)
$\alpha_1$	0.1022** (0.0333)	-0.0214 (0.0332)
$\alpha_2$	0.9050** (0.0175)	0.9639** (0.0136)
$\alpha_3$	—	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?

6. Using EViews, estimate a multivariate GARCH model for the spot and futures returns series in 'sandphedge.wf1'. Note that these series are somewhat short for multivariate GARCH model estimation. Save the fitted conditional variances and covariances, and then use these to construct the time-varying optimal hedge ratios. Compare this plot with the unconditional hedge ratio calculated in chapter 2.