Chapter 11

Modelling Variance I:
Univariate Analysis

11.1 Introduction

An important feature of many of the previous chapters is on specifying and estimating financial models of expected returns. Formally these models are based on the conditional mean of the distribution where conditioning is based on either lagged values of the dependent variable, or additional explanatory variables, or a combination of the two. From a financial perspective however, modelling the variance of financial returns is potentially more interesting because it is an important input into many aspects of financial decision making. Examples include portfolio management, the construction of hedge ratios, the pricing of options and the pricing of risk in general. In implementing these strategies, practitioners soon released that the variance, or the square root of the variance known as volatility, was time varying.

The traditional approach to modelling conditional variance is the autoregressive conditional heteroskedasticity class of models (ARCH), originally developed by Engle (1982) and extended by Bollerslev (1986) and Glosten, Jagannathan and Runkle (1993). This is a flexible class of volatility models that can capture a wide range of features that characterise time-varying risk and which generalise to multivariate settings in which time-varying models of variances and covariances are dealt with. This class of models is particularly important in modelling time-varying hedge ratios, and spillover risk.

11.2 Volatility Clustering

To investigate the econometrics of modelling time-varying volatility in both a univariate and multivariate environment, the returns on five international stock markets are investigated, namely,
SPX : Standard and Poors index from the United States;
DJX : Dow Jones index from the United States;
HSX : Hang Seng index from Hong Kong;
NKX : Nikkei index from Japan;
DAX : Deutscher Aktien Index from Germany; and
UKX : FTSE index from the United Kingdom.
The data are daily beginning 4 January 1999 and ending 2 April 2014, \( T = 3978 \). This period covers a range of crises including the dotcom bubble in early 2000, the sub-prime crisis from mid 2007 to late 2008, the Great Recession from 2008 to 2010, and the European debt crisis from 2010.

Figure 11.1: Annualised daily returns to five international stock market indices for the period 4 January 1999 to 2 April 2014 which have been standardised to have zero mean and unit variance.

One of the most documented features of financial asset returns is the tendency for large changes in asset prices to be followed by further large changes (market turmoil) or for small changes in prices to be followed by further small changes (market tranquility). This phenomenon is known as volatility clustering which highlights the property that the variance of financial returns is not constant over time, but appears to come in bursts. Figure 11.4 plots the annualised daily returns on the five international stock indices after standardisation to have zero mean and unit variance. The tendency for volatility to cluster is clearly demonstrated, particularly during the crisis periods in July of
2007 and the second half of 2008. There are also periods of tranquility when the magnitude of movements in the returns is relatively small.

A further implication of volatility clustering is that unconditional returns to the asset do not follow a normal distribution. This result is highlighted in Figure 11.2 which plots the histograms of the daily returns for each of the five stock market indices. In each case, the distribution of $r_t$ is leptokurtic, because it has a sharper peak and fatter tails than the best-fitting normal distribution, which is overlaid on the histogram in Figure 11.2.

![Histograms of daily returns for five stock market indices](image)

Figure 11.2: The distribution of the daily returns to five international stock indices over the 4 January 1999 to 2 April 2014. Superimposed on the histogram is a normal distribution with mean and variance equal to the sample mean and sample variance of the respective index returns.

To understand the relationship between volatility clustering and leptokurtosis consider a model of returns, $r_t$, which are characterised by two regimes where the variance is low in the tranquil regime and high in the turbulent regime, $h_{\text{tranquil}} < h_{\text{turbulent}}$. Assuming that the means in the two regimes are the same,

$$\mu_{\text{tranquil}} = \mu_{\text{turbulent}} = \mu$$
and that the returns in both these regimes are normally distributed, then

\[ r_t \sim \begin{cases} N\left(\mu, h_{\text{tranquil}}\right) & : \text{Tranquil regime} \\ N\left(\mu, h_{\text{turbulent}}\right) & : \text{Turbulent regime.} \end{cases} \]

The tranquil regime is characterised by returns being close to their mean \( \mu \) whereas for the turbulent regime there are large positive and negative returns which are relatively far from their mean of \( \mu \). Averaging the two distributions over the sample yields a leptokurtic distribution with the sharp peak primarily corresponding to the returns from the returns distribution during the tranquil periods. The leptokurtotic distribution is computed as the mixture distribution given by

\[ f(r) = wN\left(\mu, h_{\text{tranquil}}\right) + (1 - w)N\left(\mu, h_{\text{turbulent}}\right), \]

in which the weight \( w \) is the proportion of returns coming from each period. The parameters of the distributions in each of the regimes are estimated for the returns on the merger hedge fund index as \( \mu = 0.02, \ h_{\text{tranquil}} = 0.1, \ h_{\text{turbulent}} = 2.0 \).

There is thus a 20-fold increase in the volatility during the turbulent period, where the weight is \( w = 0.7 \) representing that 70% of returns come from the tranquil period and 30% from the period of turbulence. A plot of the two distributions is given in Figure 11.3. The fat-tails largely (if not all) corresponding to the returns from the returns distribution during the turbulent periods.

![Figure 11.3](image.png)

### 11.3 Simple Models of Time Varying Variance

By convention, the notation used for the time-varying variance in the volatility literature is \( h_t \) and not \( \sigma_t^2 \) as one might perhaps expect. This choice of \( h \)
11.3. SIMPLE MODELS OF TIME VARYING VARIANCE

is thought to be motivated by 'heteroskedasticity' which represents time-varying variance. The simplest model of time-varying variance is the historical variance, given by

$$h_t = \frac{1}{k} \sum_{i=0}^{k-1} r_{t-k}^2$$

(11.1)

where $k$ is the window over which the variance is computed and for simplicity is simply assumed that the mean return over the period may be set to zero. The advantages of this measure are that it is easy to compute involving the choice of only one parameter, namely, the window length $k$. The choice of $k$ is critical – if the it is too long then the estimate is not dynamic enough, but if it is too short the estimate will be very noisy – but the model does not offer any prescription for how $k$ is to be chosen.

The exponentially weighted moving average model (EWMA) is another simple model of time-varying variance which differs from historical volatility primarily insofar as it allows a higher weight to be attached to more recent observations. The EWMA model of variance is given by

$$h_t = (1 - \lambda) \sum_{j=1}^{\infty} \lambda^j r_{t-j}^2$$

$$= (1 - \lambda) r_{t-1}^2 + \left[ (1 - \lambda) \lambda r_{t-2}^2 + (1 - \lambda) \lambda^2 r_{t-3}^2 + \cdots \right]$$

$$= (1 - \lambda) r_{t-1}^2 + \lambda \left[ (1 - \lambda) r_{t-2}^2 + (1 - \lambda) \lambda r_{t-3}^2 + \cdots \right]$$

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

(11.2)

where $\lambda$ is known as the decay parameter which governs how recent observations are weighted relative to more distant observations. The model depends crucially on the decay parameter $\lambda$, although the model does not indicate how the crucial parameter $\lambda$ is to be estimated. In many cases a value is simply imposed with $\lambda = 0.94$ as suggested by RiskMetrics Group being a popular choice.

There are perhaps two fundamental problems with both these simple models of time-varying variance.

1. Neither model offers any prescription as to how to estimate the crucial parameters from historical data.

2. In terms of forecasting the future value of the time-varying variance, both these models suggest that the best forecast is the current estimate, $h_t$, and moreover, that this estimate is also the forecast for all future periods. This is a very undesirable feature of the models because it is to be expected that the variance will tend to revert to its long-run mean.

In order to address these fundamental flaws, an explicit dynamic model of variance is required whose parameters may be estimated from the data on historical returns.
11.4 The ARCH Model

One important implication of volatility clustering is that it should be possible to predict the evolution of the variance of returns because evidence of clustering implies autocorrelation in the variance of returns. This property may be demonstrated by realising that the square of the centered returns, $r_t^2$, provides an estimate of the variance of the returns at each $t$ and the autocorrelation function (ACF) and partial autocorrelation function (PACF) computed using $r_t^2$ should show positive and statistically significant autocorrelation. This expectation is in contrast with returns, $r_t$, which are unpredictable, at least according to the efficient markets hypothesis. Figure 11.4 plots the ACF and PACF for both the daily returns, $r_t$, and squared daily returns, $r_t^2$, to the DAX index. The results of this exercise are as expected with the autocorrelations of $r_t$ being statistically insignificant from zero, while the autocorrelations of $r_t^2$ indicate strong autocorrelation structure in $r_t^2$.

A natural test of time-variation in the variance of a variable would be to estimate an AR(p) model for the squared returns and to perform a joint test on the parameters of the lags. The idea of specifying a model that allows for time-varying variance in which an AR(p) model is estimated with the variables expressed in squares instead of in levels is now developed more formally which leads to the AutoRegressive Conditional Heteroskedasticity (ARCH) class of model introduced by Engle (1982).
To motivate the structure of the GARCH model consider the following AR(1) model of returns

\[ r_t = \phi_0 + \phi_1 r_{t-1} + u_t \]

where \( u_t \) is a disturbance term. The slope parameter \( \phi_1 \) is the first-order autocorrelation coefficient for the returns. The conditional mean given information up to time \( t - 1 \) is

\[ \mathbb{E}_{t-1}(r_t) = \phi_0 + \phi_1 r_{t-1} \]

This is the conditional mean of returns which is time-varying because it is a function of lagged returns \( r_{t-1} \).

Now consider replacing \( r_t \), by \( r_t^2 \), so the AR(1) model becomes

\[ r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + v_t \]

where \( v_t \) is another disturbance term. The slope parameter \( \alpha_1 \) is now the first-order autocorrelation coefficient of squared returns. The conditional expectation of \( r_t^2 \) given information at time \( t - 1 \) is

\[ \mathbb{E}_{t-1}(r_t^2) = \alpha_0 + \alpha_1 r_{t-1}^2. \]

Assuming that the mean of returns is zero, or that the mean has been subtracted from returns, this expression also represents the conditional variance. It is the use of lagged squared returns to model the (conditional) variance that is the key property underlying ARCH models. Moreover, the conditional variance of returns is time-varying (heteroskedastic) as it is a function of variables at time \( t - 1 \).

The ARCH model proposes a weighted average of past squared returns, similar to the historical volatility estimate as in equation (11.1), with the important improvement that the weights on the past variances are estimated from historical data. The ARCH(q) model is

\[
\begin{align*}
  r_t &= \phi_0 + \phi_1 r_{t-1} + u_t \quad \text{[Mean]} \\
  h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 \quad \text{[Variance]} \\
  u_t &\sim N(0,h_t) \quad \text{[Distribution]}
\end{align*}
\]

where the \( q \) represents the length of the lag in the conditional variance equation given by \( h_t \). The disturbance term \( u_t \) is commonly referred to as the ‘news’ because it represents the unanticipated movements in returns in excess of the conditional mean. In the special case of a constant variance \( \alpha_1 = \alpha_2 = \cdots = \alpha_q = 0 \), and the variance of \( u_t \) and hence \( y_t \), reduces to \( h_t = h = \alpha_0 \).

This observation suggests that a relatively simple test for ARCH can be performed by testing that \( \alpha_i = 0 \) for all \( i \) in a regression of the form

\[ r_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + v_t. \]
Under the null hypothesis \( E_{t-1}(r_t^2) \) will be the constant value \( \alpha_0 \). The null and alternative hypotheses are

\[
\begin{align*}
H_0 & : \alpha_i = 0 \text{ for all } i \quad \text{[No ARCH]} \\
H_1 & : \alpha_i \neq 0 \text{ for some } i \quad \text{[ARCH]}. 
\end{align*}
\]

The LM test (see Chapter 10) of these hypotheses is commonly used since it simply involves estimating an ordinary least squares regression equation and performing a goodness-of-fit test. The ARCH(\( q \)) test is implemented using the following steps.

**Step 1:** Estimate the regression equation

\[
r_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + v_t ,
\]

by ordinary least squares, where \( v_t \) is a disturbance term.

**Step 2:** Compute \( TR^2 \) from this regression and the corresponding p-value using the \( \chi^2_q \) distribution. A p-value less than 0.05 is evidence of ARCH in \( r_t \) at the 5% level.

### 11.5 The GARCH Model

A generalisation of the ARCH(\( q \)) model is the Generalised ARCH model or GARCH(p,q) model (Bollerslev, 1986), which allows for lags of the conditional variance to affect the conditional variance at time \( t \). The model is

\[
\begin{align*}
r_t &= \phi_0 + \phi_1 r_{t-1} + u_t \quad \text{[Mean]} \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \quad \text{[Variance]} \\
u_t &\sim N(0,h_t) \quad \text{[Distribution]}. 
\end{align*}
\]

Although the conditional mean is here specified to be an AR(1) model, any other specification for the mean is allowed. The key feature of the model which makes it a GARCH model is the fact that the conditional variance is given by \( q \) lags of the squared disturbance term \( u_t^2 \) and \( p \) lags of the conditional variance \( h_t \).

An important special case of this model is the GARCH(1,1) model in which the conditional variance is specified as

\[
h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} .
\]

The GARCH(1,1) model is now easily interpreted as a generalisation of the EWMA model of equation (11.2) where instead of the just one parameter (the delay parameter \( \lambda \)) there are now three unknown parameters, \( \alpha_0, \alpha_1 \) and \( \beta_1 \).

The GARCH model allows for two types of dynamics to affect the variance.
(i) Lagged shocks due to the news, \( \{u_{t-1}^2, \cdots u_{t-q}^2\} \), have a finite effect of \( q \) periods on the conditional variance \( h_t \). The effect of a shock on \( h_t \) is finite, equal to \( q \) periods.

(ii) Lagged terms in the conditional variance, \( \{h_{t-1}^2, \cdots h_{t-p}^2\} \) allow shocks to the conditional variance to have a memory longer than \( p \) periods. For example in the GARCH(1,1) model in equation (11.4), the the dynamic effects of a shock on \( h_t \) are

\[
\begin{align*}
\text{Period 1} & : \alpha_1 \\
\text{Period 2} & : \alpha_1 \beta_1 \\
& \vdots \\
\text{Period n} & : \alpha_1 \beta_1^{n-1}
\end{align*}
\]

If \( \beta_1 = 0 \), the memory is 1-period. The bigger is \( \beta_1 \), the longer is the memory of the shock.

A simple extension to the GARCH model is to allow for the effects of additional explanatory variables, \( x_1, x_2, \cdots x_K \), on the conditional moments. The GARCH model then becomes

\[
\begin{align*}
\text{Mean} & \quad r_t = \phi_0 + \phi_1 r_{t-1} + \sum_{k=1}^{K} \gamma_k x_k, t + u_t \\
\text{Variance} & \quad h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{k=1}^{K} \psi_k x_k, t \\
\text{Distribution} & \quad u_t \sim N(0, h_t)
\end{align*}
\]

Examples of potential explanatory variables are trade volumes, and dummy variables to model day-of-the-week effects and policy announcements.

### 11.6 Estimating Univariate (G)ARCH Models

GARCH models are estimated by maximum likelihood estimation which was dealt with in Chapter 10. The GARCH model in equation (11.3) specifies that the distribution of \( u_t \) is normal with zero mean and (conditional) variance \( h_t \). From this it may be deduced that the conditional distribution of \( r_t \) is

\[
f (r_t | r_{t-1}, r_{t-2}, \cdots; \theta) = \frac{1}{\sqrt{2\pi h_t}} \exp \left( -\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{2h_t} \right)
\]

Based on this conditional distribution the log-likelihood function for an observation at time \( t \) is

\[
\log L_t(\theta) = \log f (r_t | r_{t-1}, r_{t-2}, \cdots; \theta) \\
= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log h_t - \frac{1}{2} \frac{u_t^2}{h_t},
\]

(11.5)
where
\[ u_t = r_t - \phi_0 - \phi_1 r_{t-1} \]
\[ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \]
and \( \theta = \{ \phi_0, \phi_1, \alpha_1, \alpha_2, \ldots, \alpha_q, \beta_1, \beta_2, \ldots, \beta_p \} \).

To estimate the GARCH model using an iterative optimisation algorithm, a set of starting values are needed for the parameters, \( \theta_0 \), and also some initial values for computing the conditional variance. In the case of the GARCH(1,1) model the specification at observation \( t = 1 \) is
\[ h_1 = \alpha_0 + \alpha_1 u_0^2 + \beta_1 h_0 \]
so that starting values for \( u_0 \) and \( h_0 \) are required in order to compute \( h_1 \). For \( u_0 \) the mean of its distribution can be used (\( u_0 = 0 \)). For \( h_0 \) the unconditional variance can be used which is simply the sample variance of \( r_t \).

Given these starting values the evaluation of the log-likelihood function proceeds as follows.

(i) The disturbance term, \( u_t \), is evaluated for all observations using the starting values \( \theta_0 \).

(ii) Given the starting values \( \theta_0 \) and the initial values \( u_0 \) and \( h_0 \), the conditional variance \( h_t \) is evaluated recursively at all observations by using the computed values of \( u_t \) in the previous step.

(iii) Given values of \( u_t \) and \( h_t \) at each observation the log-likelihood function \( \log L(\theta_0) \) is evaluated.

The recursive computation of the log-likelihood function is embedded in a numerical optimisation routine which then solves iteratively for the maximum likelihood estimates of the parameters, \( \hat{\theta} \).

An important aspect of the estimation is that the conditional variance, \( h_t \), must always be positive at all observations both from a theoretical perspective – \( h_t \) is a variance – and from a practical perspective – the value \( \log h_t \) is computed in equation (11.5). To restrict \( h_t \) to be positive one strategy is to restrict all parameters to be positive by expressing \( h_t \) as
\[ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^2 u_{t-i}^2 + \sum_{i=1}^{p} \beta_i^2 h_{t-i} \]
Ensuring that the constraint \( h_t > 0 \) is enforced is one of the major issues faced by the various specifications of multivariate GARCH models which are introduced in Chapter 12.

The GARCH model specified so far assumes that the distribution of shocks is normal. It has already been noted that the combination of conditional normality and GARCH variance yields an unconditional distribution of financial
returns that is leptokurtotic. In practice, however, a simple GARCH model specified with normal disturbances is sometimes not able to model all of the leptokurtosis in the data. Consequently, two leptokurtotic distributions are commonly used to construct the log-likelihood function for GARCH models.

**Standardised t distribution**

Adopting the assumption that \( u_t \sim St(0, h_t, \nu) \), where \( \nu > 2 \) is the degrees of freedom parameter, implies that the conditional distribution for the GARCH(1,1) model is now

\[
f(r_t | r_{t-1}, r_{t-2}, \cdots ; \theta) = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\sqrt{\pi h_t (\nu - 2)}} \left(1 + \frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{h_t (\nu - 2)}\right)^{-(\nu + 1) / 2}
\]

where \( \theta = \{ \phi_0, \phi_1, \alpha_1, \alpha_2, \cdots, \alpha_q, \beta_1, \beta_2, \cdots, \beta_p, \nu \} \). The log-likelihood function for observation \( t \) is

\[
\log L_t(\theta) = -\frac{1}{2} \log (\pi (\nu - 2)) - \frac{1}{2} \log (h_t) + \log \left( \Gamma \left(\frac{\nu + 1}{2}\right)\right) - \log \left( \Gamma \left(\frac{\nu}{2}\right)\right)
\]

\[
- \left(\frac{\nu + 1}{2}\right) \log \left(1 + \frac{(u_t - \phi_0 - \phi_1 r_{t-1})^2}{h_t (\nu - 2)}\right)
\]

with

\[
u_t = r_t - \phi_0 - \phi_1 r_{t-1}
\]

\[
h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}.
\]

As before, an iterative optimisation algorithm is needed to estimate the parameters of the model by maximum likelihood.

The degrees of freedom parameter \( \nu \) must be constrained to be positive and greater in value than 2. For \( \nu < 2 \) the variance of the t distribution does not exist in which case trying estimate the conditional variance is counterintuitive.

**Generalised Error Distribution (ged)**

Adopting the assumption that \( u_t \sim ged(0, h_t, s) \), where \( s \) is the shape parameter, implies that the conditional distribution for the GARCH(1,1) model is now

\[
f(r_t | r_{t-1}, r_{t-2}, \cdots ; \theta) = \frac{s \exp \left( -\frac{1}{2} \frac{|r_t - \phi_0 - \phi_1 r_{t-1}|^5}{\lambda \sqrt{h_t}} \right)}{\lambda^{(1 + 1/s)}} \frac{1}{\Gamma(\frac{1}{s})}
\]

in which \( s > 0 \) and

\[
\lambda = \left[ \frac{\Gamma(1/\nu)}{2^{2/\nu} \Gamma(3/\nu)} \right]^{1/2}.
\]
Table 11.1

Parameter estimates of GARCH(1,1) models for the daily returns to five international stock indices for log-likelihood functions based on the normal, $t$ and generalised error distributions. The sample period is 4 January 1999 to 2 April 2014.

<table>
<thead>
<tr>
<th></th>
<th>SPX</th>
<th>DJX</th>
<th>HSX</th>
<th>NKX</th>
<th>DAX</th>
<th>UKX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>0.045</td>
<td>0.050</td>
<td>0.046</td>
<td>0.044</td>
<td>0.073</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.018)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.014</td>
<td>0.013</td>
<td>0.012</td>
<td>0.041</td>
<td>0.024</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>0.077</td>
<td>0.082</td>
<td>0.056</td>
<td>0.090</td>
<td>0.086</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.913</td>
<td>0.907</td>
<td>0.939</td>
<td>0.893</td>
<td>0.903</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>0.059</td>
<td>0.057</td>
<td>0.051</td>
<td>0.053</td>
<td>0.085</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.010</td>
<td>0.010</td>
<td>0.008</td>
<td>0.030</td>
<td>0.016</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>0.079</td>
<td>0.082</td>
<td>0.050</td>
<td>0.074</td>
<td>0.084</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.917</td>
<td>0.913</td>
<td>0.948</td>
<td>0.914</td>
<td>0.912</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td></td>
<td>(0.730)</td>
<td>(0.710)</td>
<td>(0.711)</td>
<td>(0.893)</td>
<td>(1.072)</td>
<td>(1.383)</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>0.051</td>
<td>0.043</td>
<td>0.021</td>
<td>0.019</td>
<td>0.083</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td>(0.018)</td>
<td>(0.016)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.011</td>
<td>0.011</td>
<td>0.008</td>
<td>0.035</td>
<td>0.019</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.010)</td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>0.079</td>
<td>0.083</td>
<td>0.052</td>
<td>0.079</td>
<td>0.085</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.915</td>
<td>0.911</td>
<td>0.946</td>
<td>0.907</td>
<td>0.908</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>$s$</td>
<td>1.277</td>
<td>1.270</td>
<td>1.172</td>
<td>1.258</td>
<td>1.391</td>
<td>1.464</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.035)</td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.041)</td>
<td>(0.045)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
The log-likelihood function for observation $t$ is

$$
\log L_t(\theta) = \log \left( \frac{s}{\lambda} \right) - \left( 1 + \frac{1}{s} \right) \log 2 - \log \Gamma \left( \frac{1}{s} \right) - \frac{1}{2} \log h_t - \frac{1}{2} \left( \frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\lambda^2 h_t} \right)^{\frac{1}{s}}
$$

The increasing use of the ged distribution in estimating GARCH models derives from its versatility. In particular

- $s > 2$ ged has thinner tails than the normal distribution
- $s = 2$ ged is identical to the normal distribution
- $s < 2$ ged has fatter tails than the normal distribution

While the ged is capable of generating very fat tails it cannot match the $t$ distribution. When $\nu < 4$ in the $t$ distribution the the tails are so fat that the kurtosis is infinite. So in general for financial data, it is to be expected that parameter values of $\nu > 4$ will result when using the $t$ distribution, while the ged distribution should yield estimates $s < 2$. A GARCH(1,1) model is fitted to the the six international stock index returns for the period 4 January 1999 to 2 April 2014 using each of the three distributions discussed in this section. The results are reported in Table 11.1.

An interesting feature of the empirical results is the consistency of the parameter estimates across all five equity returns and also across the three distributions. The estimates of $\beta_1$ are all in the vicinity of 0.9 and the estimates of $\alpha_1$ being between 0.05 and 0.095. A pleasing result is that the parameters $\nu u$ and $s$ are both in the expected range and consistently indicate that the normal distribution is perhaps not appropriate in these examples. On the other hand, the similarity of the estimated parameters suggests that choice of appropriate distribution will need to be made on the basis of features other than point estimates of the parameters, such as forecast performance or the distribution of the standardised residuals.

Very often in GARCH(1,1) models of equity returns $\hat{\alpha}_1 + \hat{\beta}_1 \simeq 1$. The sum $\hat{\alpha}_1 + \hat{\beta}_1$ for these indices appears to be in keeping with this general observation, although the sum in each case is less than one. If $\hat{\alpha}_1 + \hat{\beta}_1 = 1$ then the volatility series is nonstationary, a situation known as IGARCH, where the ‘I’ stands for integrated following the discussion of nonstationary models in Part ??.

A simpler approach to dealing with the problem of leptokurtosis is to recognise that the assumption of normally distributed disturbances results in the misspecification of the log-likelihood function. Despite the shape of the distribution being incorrect, the mean and variance of this distribution are, however, correctly specified. It turns out that estimation of the parameters of the conditional mean and the conditional variance are still consistent but the standard errors require correction. The corrected standard errors of the maximum likelihood estimators of $\theta$ that are consistent are known as Bollerslev-Wooldridge standard errors (Bollerslev and Wooldridge, 1992). Computation of these standard errors involves using information in both the Hessian matrix and the outer product of gradients matrix.
11.7 Asymmetric Volatility Effects

Consider the GARCH(1,1) model

\[
\begin{align*}
    r_t &= \phi_0 + u_t \\
    h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}.
\end{align*}
\]

For no news-days, \( u_{t-1} = 0 \) the conditional variance has a minimum value at \( h_t = \alpha_0 \). An important property of this GARCH(1,1) specification is that shocks of the same magnitude, positive or negative, result in the same increase in volatility \( h_t \). That is, positive news, \( u_{t-1} > 0 \), has the same effect on the conditional variance as negative news \( u_{t-1} < 0 \) because it is only the absolute size of the news that matters since \( u_{t-1}^2 \) enters the equation. In the case of stock markets, an asymmetric response to the news in which negative shocks \( u_{t-1} < 0 \) have a larger effect on conditional variance is supported by theory. A negative shock raises the debt-equity ratio, thereby increasing leverage and consequently risk and this so-called leverage effect therefore suggests that bad news causes a greater increase in conditional variance than good news.

There are two popular specifications in the GARCH class of model that relax the restriction of a symmetric response to the news.

1. **Threshold GARCH (TGARCH):**

   The TGARCH specification of the conditional variance is

   \[
   h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1} I_{t-1}
   \]

   where \( I_{t-1} \) is an indicator variable defined as

   \[
   I_{t-1} = \begin{cases} 
   1 & : \ u_{t-1} \geq 0 \\
   0 & : \ u_{t-1} < 0
   \end{cases}
   \]

   To make the asymmetry in the effect of news on the conditional variance explicit, this model can also be written as

   \[
   h_t = \begin{cases} 
   \alpha_0 + (\alpha_1 + \lambda) u_{t-1}^2 + \beta_1 h_{t-1} & : \ u_{t-1} \geq 0 \\
   \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} & : \ u_{t-1} < 0
   \end{cases}
   \]

   If \( \lambda > 0 \) then positive news, \( u_{t-1} \geq 0 \), has a greater effect on volatility than negative news. The leverage effect in equity markets would lead us to expect \( \lambda < 0 \) so that negative news, \( u_{t-1} < 0 \), is associated with a higher effect on volatility than positive news.

2. **Exponential GARCH (EGARCH):**

   The EGARCH specification of the conditional variance is

   \[
   \log h_t = \alpha_0 + \sum_{i=1}^{q} \left( \alpha_i \left( \frac{u_{t-i}}{\sqrt{h_{t-i}}} \right) + \lambda_i \frac{u_{t-i}}{\sqrt{h_{t-i}}} \right) + \sum_{j=1}^{p} \beta_j \log(h_{t-j})
   \]
11.7. ASYMMETRIC VOLATILITY EFFECTS

An important advantage of the EGARCH specification is that the conditional variance is guaranteed to be positive at each point in time because the variance is expressed in terms of log $h_t$ so that the actual variance obtained by exponentiation.

ARCH(1), GARCH(1,1), TARCH(1,1) and EGARCH(1,1) models were fitted to daily returns from 4 January 1999 to 2 April 2014 on the returns to six international stock indices. The results are reported in Table 11.2.

Table 11.2

| Parameter estimates of TARCH(1,1) models for the daily returns to five international stock indices expressed as percentages. The sample period is 4 January 1999 to 2 April 2014. |
|---|---|---|---|---|---|
| Mean | SPX | DJX | HSX | NKX | DAX | UKX |
| $\phi_0$ | 0.003 | 0.013 | 0.024 | 0.015 | 0.026 | -0.001 |
| (0.014) | (0.013) | (0.018) | (0.020) | (0.018) | (0.014) |
| Variance | $\alpha_0$ | 0.014 | 0.013 | 0.016 | 0.053 | 0.029 | 0.017 |
| (0.001) | (0.001) | (0.003) | (0.008) | (0.003) | (0.002) |
| $\alpha_1$ | 0.129 | 0.137 | 0.081 | 0.131 | 0.139 | 0.136 |
| (0.009) | (0.010) | (0.006) | (0.010) | (0.010) | (0.009) |
| $\beta_1$ | 0.938 | 0.926 | 0.939 | 0.889 | 0.911 | 0.920 |
| (0.005) | (0.006) | (0.005) | (0.008) | (0.007) | (0.007) |
| $\gamma$ | -0.157 | -0.149 | -0.057 | -0.090 | -0.133 | -0.143 |
| (0.010) | (0.010) | (0.006) | (0.008) | (0.009) | (0.010) |

As expected $\lambda < 0$ for each of the returns series considered and the parameter is statistically significant indicating the presence of the leverage effect in these markets. A plot of $h_t$ against $u_t - 1$, is known as the news impact curve. The news impact curve illustrates quite sharply the differences between the various specifications of the conditional variance. To demonstrate this point an ARCH(1), GARCH(1,1), TARCH(1,1) and EGARCH(1,1) model is fitted to the returns to the S&P 500 and for each model the news impact curve is plotted in Figure ??.

The major point to note is that the simple ARCH and GARCH models impose a symmetric news impact curve whereas the TARCH and EGARCH models relax this assumption and allow for asymmetric adjustment to the news. For the models estimated here, the news impact curve is much flatter for positive shocks than it is for negative shocks, indicating that negative news has a much larger impact on volatility than positive news. The situation depicted here of the the news impact curve actually decreasing as positive shocks get larger is not typical of many applications for stock market returns.
11.8 The Risk-Return Trade-off

The standard deviation is commonly used as a measure of risk in portfolio theory as it represents the deviations of actual returns from their conditional mean. The larger the deviation the larger is the risk of the portfolio. To compensate an investor for bearing more risk, an investor should receive a higher expected return, resulting in a positive relationship between the mean and the risk of the portfolio.

Letting \( \mu_t = \mathbb{E}_{t-1}[r_t] \) represent the conditional mean of the portfolio and \( \mathbb{E}_{t-1}[r_t^2] = h_t \) represent the conditional variance of the portfolio. The fundamental relationship in finance between risk and return is specified as

\[
\mu_t = \phi_0 + \phi_1 h_t^{\omega} = \phi_0 + \phi_1 \sigma_t^{2\omega}
\]

where, for notational convenience the conditional standard deviation or risk of the portfolio is denoted \( \sqrt{h_t} = \sigma_t \) and \( \phi_1 > 0 \) to allow for a positive relationship between the expected return and risk.

The compensation an investor wishes to receive from bearing higher risk is
11.8. THE RISK-RETURN TRADE-OFF

given by
\[ \frac{d \mu_t}{d \sigma_t} = 2 \omega \phi_1 \sigma_t^2 \omega - 1, \]
giving rise to two special cases.

(i) **Case 1:** $\omega = 0.5$
There is a linear relationship between the mean, $\mu_t$, and the conditional standard deviation of the portfolio, $\sigma_t$. Compensation for bearing more risk increases at the constant rate given by
\[ \frac{d \mu_t}{d \sigma_t} = \phi_1. \]

(ii) **Case 2:** $\omega = 1.0$
There is a nonlinear relationship between the mean, $\mu_t$, and the conditional standard deviation of the portfolio, $\sigma_t$. Compensation for bearing more risk (increase in $\sigma_t$) increases at an increasing rate
\[ \frac{d \mu_t}{d \sigma_t} = 2 \phi_1 \sigma_t \]

To illustrate the risk-return trade-off data for the 10 Fama-French industry portfolios, which comprises monthly data beginning January 1927 and ending December 2013, are used. Define the excess returns to the portfolios and the market, respectively as
\[ z_{it} = r_{it} - r_{ft}, \quad z_{mt} = r_{mt} - r_{ft}, \]
The then the relationship between risk and return may be captured in the GARCH(1,1) framework by specifying the conditional mean as a function of the conditional variance as follows
\[ z_{it} = \phi_0 + \phi_1 h_t \omega + \phi_2 z_{mt} + u_t \]
\[ u_t \sim N(0, h_t) \] (11.6)
\[ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}. \]
The model is an extension of the CAPM with an allowance for time-varying risk to identify the trade-off preferences of investor between risk and return and the specification is known as GARCH-M.
The critical parameters reflecting the risk preferences of investors are $\phi_1$ and $\omega$. The GARCH-M augmented version of the CAPM is estimated for 10 industry portfolios using values $\omega = \{0.5, 1.0\}$ and Table 11.3 summarises the results relating to the trade-off parameter $\phi_1$ for all 10 portfolios. Inspection of the results for $\omega = 0.5$ shows that the Nondurables, Durables and Retail portfolios exhibit the greatest trade-off, all with $\hat{\phi}_1 \simeq 0.3$. The Utilities and
Manufacturing portfolio have the smallest positive trade-offs, while the Energy and Health portfolios even exhibit a negative trade-off. The issue of a negative trade-off is investigated further below by testing the strength of the risk-return relationships.

Table 11.3

Estimation of parameter $\phi_1$ the GARCH-M version of the CAPM in (11.6) with $\omega = \{0.5, 1.0\}$ for 10 Fama-French industry portfolios. The data are monthly excess returns for the period January 1927 to December 2013.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$\phi_1$</th>
<th>t test</th>
<th>p value</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nondurables</td>
<td>0.286</td>
<td>1.859</td>
<td>0.063</td>
<td>4468.639</td>
</tr>
<tr>
<td>Durables</td>
<td>0.378</td>
<td>2.861</td>
<td>0.004</td>
<td>5718.193</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.135</td>
<td>1.068</td>
<td>0.285</td>
<td>3925.937</td>
</tr>
<tr>
<td>Energy</td>
<td>-0.053</td>
<td>-0.430</td>
<td>0.667</td>
<td>5680.519</td>
</tr>
<tr>
<td>Technology</td>
<td>0.052</td>
<td>0.357</td>
<td>0.721</td>
<td>5188.485</td>
</tr>
<tr>
<td>Telecom.</td>
<td>0.131</td>
<td>0.931</td>
<td>0.352</td>
<td>5156.026</td>
</tr>
<tr>
<td>Retail</td>
<td>0.301</td>
<td>1.886</td>
<td>0.059</td>
<td>4999.628</td>
</tr>
<tr>
<td>Health</td>
<td>-0.276</td>
<td>-1.881</td>
<td>0.060</td>
<td>5388.758</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.011</td>
<td>0.120</td>
<td>0.904</td>
<td>5407.391</td>
</tr>
<tr>
<td>Other</td>
<td>0.115</td>
<td>1.170</td>
<td>0.242</td>
<td>4362.822</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$\phi_1$</th>
<th>t test</th>
<th>p value</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nondurables</td>
<td>0.064</td>
<td>2.053</td>
<td>0.040</td>
<td>4467.797</td>
</tr>
<tr>
<td>Durables</td>
<td>0.039</td>
<td>2.403</td>
<td>0.016</td>
<td>5718.994</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.044</td>
<td>1.299</td>
<td>0.194</td>
<td>3924.829</td>
</tr>
<tr>
<td>Energy</td>
<td>-0.010</td>
<td>-0.599</td>
<td>0.549</td>
<td>5680.353</td>
</tr>
<tr>
<td>Technology</td>
<td>0.001</td>
<td>0.062</td>
<td>0.950</td>
<td>5188.598</td>
</tr>
<tr>
<td>Telecom.</td>
<td>0.016</td>
<td>0.741</td>
<td>0.459</td>
<td>5156.408</td>
</tr>
<tr>
<td>Retail</td>
<td>0.055</td>
<td>1.964</td>
<td>0.050</td>
<td>4998.995</td>
</tr>
<tr>
<td>Health</td>
<td>-0.029</td>
<td>-1.504</td>
<td>0.133</td>
<td>5390.037</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.000</td>
<td>0.006</td>
<td>0.995</td>
<td>5407.404</td>
</tr>
<tr>
<td>Other</td>
<td>0.037</td>
<td>2.170</td>
<td>0.030</td>
<td>4360.769</td>
</tr>
</tbody>
</table>

A test of a trade-off between risk and return is based on the hypotheses

$H_0 : \phi_1 = 0$ [No tradeoff]

$H_1 : \phi_1 \neq 0$ [Tradeoff]

The Wald test of these hypotheses is conveniently given by the $t$ statistic. The $p$ values of the models when $\omega = 0.5$ indicate that only the Durables portfolio has a statistically significant trade-off between risk and return at the conventional level of 5%. The results for the Nondurables and Retail portfolios are marginal, as is the result for the Health portfolio which has the counterintuitive negative relationship. The situation is more promising for the ability of
11.9. FORECASTING

this model to capture the risk-return relationship when $\omega = 1.0$. The Non-
durables, Durables, Retail and Other portfolios are all indicate a significant
risk-return tradeoff and the anomalous result for the Health portfolio is re-
solved because $\hat{\phi}_1$ is not significant.
Table 11.3 also reports the Akaike Information Criteria (AIC) which may be
used as a test to determine what type of risk preferences are most consistent
with the portfolios because the models based on $\omega = 0.5$ and $\omega = 1.0$, are
nonnested. As discussed in Chapter 4, the AIC statistic is computed as

$$AIC = -2\log L(\hat{\theta}) + \frac{2K}{T}$$

where $K = 7$ is the number of estimated parameters which applies to both
models.
A comparison of the AICs for the two models associated with all 10 portfo-
lios reveals an even split between the portfolios for which the statistic is min-
imised when $\omega = 0.5$ (Nondurables, Technology, Telecom., Health and Utili-
ties) and $\omega = 1.0$ (Durables, Manufacturing, Energy, Retail and Other).

11.9. Forecasting

Forecasting GARCH models is similar to forecasting ARMA models discussed
in Chapter 7. The only difference is that with ARMA forecasts the focus is on
the level of the series whereas with GARCH forecasts it is on the variance of
the series. To highlight the process of forecasting GARCH conditional vari-
ances, consider the GARCH(1,1) model. To forecast volatility at time $T+1$,
the conditional variance is written at $T+1$

$$h_{T+1} = a_0 + \alpha_1 u_T^2 + \beta_1 h_T$$

Taking conditional expectations based on information at time $T$, the one-step
ahead forecast of $h_T$ is

$$h_{T+1|T} = \mathbb{E}_T [h_{T+1}] = \mathbb{E}_T \left[a_0 + \alpha_1 u_T^2 + \beta_1 h_T\right]$$

$$= a_0 + \alpha_1 u_T^2 + \beta_1 h_T$$

since $\mathbb{E}_T [u_T^2] = u_T^2$ and $\mathbb{E}_T [h_T] = h_T$. Similarly, to forecast volatility at time
$T+2$, the conditional variance is written at $T+2$

$$h_{T+2} = a_0 + \alpha_1 u_{T+1}^2 + \beta_1 h_{T+1}$$

Taking conditional expectations based on information at time $T$, the two-step
ahead forecast of $h_{T+2}$ is

$$h_{T+2|T} = \mathbb{E}_T [h_{T+2}] = \mathbb{E}_T \left[a_0 + \alpha_1 u_{T+1}^2 + \beta_1 h_{T+1}\right]$$

$$= a_0 + \alpha_1 \mathbb{E}_T [u_{T+1}^2] + \beta_1 \mathbb{E}_T [h_{T+1}]$$

$$= a_0 + \alpha_1 h_{T+1|T} + \beta_1 h_{T+1|T}$$

$$= a_0 + \left(\alpha_1 + \beta_1\right) h_{T+1|T}$$

(11.7)
as by definition $\mathbb{E}_T[u_{T+1}^2] = h_{T+1|T}$ and $\mathbb{E}_T[h_{T+1}] = h_{T+1|T}$. By extending this argument to $T+k$ gives

$$h_{T+k|T} = a_0 + (a_1 + \beta_1) h_{T+k-1|T}$$

Recursive substitution for the term $h_{T+k-1|T}$ in (11.8), using the results of the form (11.7), the conditional forecast of volatility for $k$ periods ahead is

$$h_{T|T} = a_0 + (a_1 + \beta_1) a_0 + \cdots + (a_1 + \beta_1)^{k-2} a_0 + (a_1 + \beta_1)^{k-1} h_{T+1|T}$$

In summary, the forecasts for the GARCH(1,1) model are summarised as

$$h_{T+1|T} = a_0 + a_1 u_T^2 + \beta_1 h_T$$
$$h_{T+k|T} = a_0 + (a_1 + \beta_1) h_{T+k-1|T}, \quad k \geq 2$$

In practice, these forecasts for the GARCH(1,1) model are computed by replacing the unknown parameters $a_0$, $a_1$ and $\beta_1$ and the unknown quantities $u_T^2$ and $h_T$ by their respective sample estimates. The forecasts are computed recursively staring with

$$\hat{h}_{T+1|T} = \hat{a}_0 + \hat{a}_1 \hat{u}_T^2 + \hat{\beta}_1 \hat{h}_T$$

Given this estimate, $\hat{h}_{T+2|T}$ is computed from (11.7) as

$$\hat{h}_{T+2|T} = \hat{a}_0 + (\hat{a}_1 + \hat{\beta}_1) \hat{h}_{T+k-1|T}$$

which, in turn, is used to compute $\hat{h}_{T+3|T}$ etc. To forecast higher order GARCH models the same recursive approach is adopted.

One of the main issues highlighted in Section ?? with forecasting time-varying variances using either a historical average or an exponentially weighted moving average, is that the current forecast of the variance is also the expected future value of the variance. In other words, if variance is at historically high levels when the EWMA estimate is computed then this high value for the variance is forecast to continue indefinitely. By contrast, the forecast from a GARCH(1,1) model will converge relatively quickly to the long-term average volatility implied by the model, which is given by

$$\overline{h} = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$
above that the long-term mean for the July forecast. The fact that the convergence occurs over a 12 month period indicates that the conditional volatility series is quite persistent. Notice that for the forecast starting in July 2010, the actual estimated conditional variance series drops off a lot more quickly than the forecast.

One of the distinguishing features of the conditional variance literature has been the rapid proliferation of types of model available. Factors which have to be chosen range from the specification of the mean process through the choice of specification for the conditional variance to the selection of the appropriate error distribution on which to base the construction of the log-likelihood function. Given this overwhelming choice, one of the more interesting results to emerge is that despite its simplicity, when it comes to forecasting the conditional variance, the simple GARCH(1,1) model is difficult to beat (Hansen and Lunde, 2005).

The claimed efficacy of the GARCH(1,1) for forecasting conditional variance then naturally leads to the question of assessing the accuracy of variance forecasts. In theory, determining the accuracy of the forecasts of the conditional variance can be accomplished using any of the statistical measures outlined in Chapter 7. In practice, however, this proves difficult because it is possible to compare the forecasts with the actual value of the conditional variance with its forecast because the former is never directly observed.

The standard method to assess volatility models, therefore, is to evaluate the forecast using a volatility proxy and such as the squared return, $r_t^2$. Early at-
tempts are forecast evaluation were based on Mincer-Zarnowitz regressions (Mincer and Zarnowitz, 1969) in which the realisation of the variable of interest is regressed on the forecast

\[ r_t^2 = \delta_0 + \delta_1 \hat{h}_t + u_t. \]

The null and alternative hypotheses are

\[ H_0 : \delta_0 = 0 \text{ and } \delta_1 = 1 \]
\[ H_1 : \delta_0 \neq 0 \text{ or } \delta_1 \neq 1. \]

The use of \( r_t^2 \) as a proxy is problematic, however, as returns that are large in absolute value may have a large impact on the estimation results. Two examples of alternative specifications that have been tried are

\[ |r_t| = \delta_0 + \delta_1 \sqrt{\hat{h}_t} + u_t \]
\[ \log r_t^2 = \delta_0 + \delta_1 \log \hat{h}_t + u_t, \]

which use transformations of the volatility proxy to reduce the impact of large returns.

In general, the trend has been to move away from Mincer-Zarnowitz regressions and to use the measures of forecast performance outlined in Chapter 7 together with the Diebold-Mariano test (Diebold and Mariano, 1995) to assess volatility forecasts. This has been a particularly fertile area of research and has seen the development of new loss functions, such as the quasi-likelihood loss function, which is defined for observation \( t \) as

\[ QLIKE = \log \hat{h}_t + \frac{r_t^2}{\hat{h}_t}. \]

The name QLIKE is derived from the similarity to the (negative) Gaussian log-likelihood and its use as a quasi-likelihood in mis-specified models. Specified in this way the QLIKE function can become negative when dealing with very small returns because the term in \( \log \hat{h}_t \) will be negative and dominate the other term in the expression. To avoid this, an equivalent alternative specification (see Christoffersen, 2012) which is always positive is

\[ QLIKE = \frac{r_t^2}{\hat{h}_t} - \log \left( \frac{r_t^2}{\hat{h}_t} \right) - 1. \]

The QLIKE function has become very popular in evaluating variance forecasts. The major reason for this popularity is the fact that the QLIKE criterion is that it is not symmetric. Figure 11.7 plots the RMSE and the QLIKE measures for forecasts ranging from 0.5 to 3 when the true value is 2. Unlike the RMSE, the QLIKE penalises underestimating the volatility more heavily than overestimating it. This may be a desirable characteristic in a loss function if
Figure 11.7: The RMSE (dashed line) and QLIKE loss functions plotted for a forecasts ranging from 0.5 to 3.5 for the true value 2.

Table 11.4

<table>
<thead>
<tr>
<th>Criterion</th>
<th>GARCH(1,1)</th>
<th>EWMA</th>
<th>DM t test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.000</td>
<td>0.000</td>
<td>−17.895</td>
<td>0.000</td>
</tr>
<tr>
<td>MAPE</td>
<td>1982.110</td>
<td>2946.970</td>
<td>−1.464</td>
<td>0.144</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.000</td>
<td>0.000</td>
<td>−3.460</td>
<td>0.001</td>
</tr>
<tr>
<td>QLIKE</td>
<td>1.986</td>
<td>2.090</td>
<td>−1.968</td>
<td>0.049</td>
</tr>
</tbody>
</table>

the risk manager is particularly conservative. Other loss functions based on economic notions of loss such as expected utility have also been proposed.

Table 11.4 provides a comparison of the forecasts generated by a GARCH(1,1) model fitted to S&P 500 returns against a forecast of the conditional variance using an exponentially weighted moving average of squared returns. The forecasts start at 1 July 2010 and are made for 500 days after this date. The EWMA forecast is computed using the weighting parameter 0.94 as suggested by RiskMetrics. The MAE, MAPE and RMSE as discussed in Chapter 7 are reported together with the Diebold-Mariano test of equal predictive accuracy. The QLIKE metic and associated Diebold-Mariano test are also reported. It is quite clear from these results that the GARCH(1,1) model dominates the EWMA in terms of forecasting accuracy. It is only in the MAPE case that the Diebold-Mariano test of equal predictive ability cannot be rejected at the 5% level and in each case the t statistic is negative indicating that the GARCH(1,1) loss function has a smaller value than the EWMA loss function.
11.10 Exercises

1. Time-variation in Hedge Funds

This question is based on the EViews file HEDGE.WF1 which contains daily data on the percentage returns of seven hedge fund indexes, from the 1st of April 2003 to the 28th of May 2010, a sample size of \( T = 1869 \).

   \[
   \begin{align*}
   \text{R_CONVERTIBLE} & : \text{Convertible Arbitrage} \\ 
   \text{R_DISTRESSED} & : \text{Distressed Securities} \\ 
   \text{R_EQUITY} & : \text{Equity Hedge} \\ 
   \text{R_EVENT} & : \text{Event Driven} \\ 
   \text{R_MACRO} & : \text{Macro} \\ 
   \text{R_MERGER} & : \text{Merger Arbitrage} \\ 
   \text{R_NEUTRAL} & : \text{Equity Market Neutral}
   \end{align*}
   \]

   (a) Using the returns on the Merger hedge fund estimate the constant mean model

   \[ R_{\text{MERGER}} t = \gamma_0 + u_t, \]

   and interpret the time series properties of \( \hat{u}_t \) and \( \hat{u}_t^2 \), where \( \hat{u}_t \) is the demeaned return.

   (b) Compute the empirical distribution of \( \hat{u}_t \). Perform a test of normality and interpret the result.

   (c) Test for ARCH of orders \( p = 1, 2, 5, 10 \), in the Merger hedge fund returns.

   (d) Repeat parts (a) and (b) for the other six hedge funds.

2. Time-variation in Stock Market Indexes

This question is based on the EViews file HEDGE.WF1 which contains daily data on percentage returns of the S&P500, DOW and NASDAQ, from the 1st of April 2003 to the 28th of May 2010, a sample size of \( T = 1869 \).

   (a) Using the returns on the S&P500 index estimate the constant mean model

   \[ R_{\text{SP500}} t = \gamma_0 + u_t, \]

   and interpret the time series properties of \( \hat{u}_t \) and \( \hat{u}_t^2 \), where \( \hat{u}_t \) is the demeaned return.

   (b) Compute the empirical distribution of \( \hat{u}_t \). Perform a test of normality and interpret the result.

   (c) Test for ARCH of orders \( p = 1, 2, 5, 10 \), in the S&P500 returns.

   (d) Repeat parts (a) and (b) for the DOW and NASDAQ stock market indexes.
3. **GARCH Models of Hedge Funds**

This question is based on the EViews file `HEDGE.WF1` used in Question 1.

(a) Using the returns on the Merger hedge fund estimate the GARCH(1,1) model

\[
R_{MERGER_t} = \gamma_0 + u_t \\
\quad u_t \sim N(0, h_t) \\
h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}.
\]

Sketch the news impact curve.

(b) Extend part (a) by estimating the TARCH(1,1) model

\[
R_{MERGER_t} = \gamma_0 + u_t \\
\quad u_t \sim N(0, h_t) \\
h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}d_{t-1}.
\]

Sketch the news impact curve. The parameter \(\lambda\) is commonly interpreted as the leverage effect in models of equity returns. Perform a test of symmetry by testing the restriction \(\lambda = 0\). Interpret the results of the test.

(c) Expand the TARCH model in part (b) to allow for day-of-the-week effects in both the conditional mean and the conditional variance. Perform a test of these effects in the

i. Conditional mean
ii. Conditional variance
iii. Conditional mean and the conditional variance.

(d) Given the results of part (c), reestimate the model by replacing the conditional normality distribution by the conditional standardized Student t distribution. Interpret the "degrees of freedom" parameter estimate.

(e) Repeat parts (a) to (d) for the other size hedge funds. For each model compare the parameter estimates associated with each of the seven hedge funds.

4. **GARCH Models of Stock Market Indexes**

This question is based on the EViews file `HEDGE.WF1` used in Question 2.

(a) Using the daily percentage returns on the S&P500 index, estimate the GARCH(1,1) model

\[
R_{SP500_t} = \gamma_0 + u_t \\
\quad u_t \sim N(0, h_t) \\
h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}.
\]
Sketch the news impact curve.

(b) Extend part (a) by estimating the TARCH(1,1) model and sketch the news impact curve

\[
R_{SP500t} = \gamma_0 + u_t \\
u_t \sim N(0, h_t) \\
h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}^2 d_{t-1}.
\]

Sketch the news impact curve. Perform a test of symmetry by testing the restriction $\lambda = 0$. Interpret the results of the test.

(c) Expand the TARCH model in part (b) to allow for day-of-the-week effects in both the conditional mean and the conditional variance. Perform a test of these effects in the

i. Conditional mean
ii. Conditional variance
iii. Conditional mean and the conditional variance.

(d) Given the results of part (c), reestimate the model by replacing the conditional normality distribution by the conditional Standardized Student t. Interpret the “degrees of freedom” parameter estimate.

(e) Repeat parts (a) to (d) for the DOW and the NASDAQ. For each model compare the parameter estimates associated with each of the three stock market indexex.

(f) Suggest an alternative conditional volatility model to the TARCH model that allows for shocks to have asymmetric effects on volatility. Briefly discuss the advantages of the two alternative specifications.

5. Time-varying Risk in Hedge Funds

This question is based on the EViews file HEDGE.WF1 used in Question 1.

(a) Estimate the following model

\[
R_{MERGERt} = \gamma_0 + \theta h_t^\beta + u_t \\
u_t \sim N(0, h_t) \\
h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}^2 d_{t-1},
\]

with $\beta_1 = 1$. Test for a time-varying risk premium by testing the restriction $\theta = 0$. Interpret the result.

(b) Repeat part (a) with $\beta = 0.5$. 
(c) Estimate the following model

\[ R_{\text{MERGER}}_t = \gamma_0 + \theta \log h_t + u_t \]
\[ u_t \sim N(0, h_t) \]
\[ h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}^2 d_{t-1}. \]

Test for a time-varying risk premium by testing the restriction \( \theta = 0 \). Interpret the result.

(d) Briefly discuss the differences in the risk preferences associated with the models in parts (a) to (c).

(e) Repeat parts (a) to (d) for the other six hedge funds. Compare the estimates of time-varying risk premium of the seven indexes.

6. Time-varying Risk in Stock Market Indexes

This question is based on the EViews file HEDGE.WF1 used in Question 2.

(a) Estimate the following model

\[ R_{\text{SP500}}_t = \gamma_0 + \theta h_t^\beta + u_t \]
\[ u_t \sim N(0, h_t) \]
\[ h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}^2 d_{t-1}. \]

with for \( \beta = 1 \). Test for a time-varying risk premium by testing the restriction \( \theta = 0 \). Interpret the result.

(b) Repeat part (a) with \( \beta = 0.5 \).

(c) Estimate the following model

\[ R_{\text{SP500}}_t = \gamma_0 + \theta \log h_t + u_t \]
\[ u_t \sim N(0, h_t) \]
\[ h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}^2 d_{t-1}. \]

Test for a time-varying risk premium by testing the restriction \( \theta = 0 \). Interpret the result.

(d) Briefly discuss the differences in the risk preferences associated with the models in parts (a) to (c).

(e) Repeat parts (a) to (d) for the DOW and NASDAQ. Compare the estimates of time-varying risk premium of the three indexes.

7. Capital Asset Pricing Model of Hedge Funds

This question is based on the EViews file HEDGE.WF1 used in Question 1.
CHAPTER 11. MODELLING VARIANCE I: UNIVARIATE ANALYSIS

(a) Estimate the following CAPM with constant variance for the Merger hedge fund

\[ R_{\text{MERGER}}_t = \gamma_0 + \gamma_1 R_{\text{SP500}}_t + u_t \]
\[ u_t \sim N(0, h_t) \]
\[ h_t = \alpha_0. \]

Interpret the parameter estimates and compute estimates of the idiosyncratic risk and the systematic risk.

(b) Estimate the following CAPM with time-varying variance for the Merger hedge fund

\[ R_{\text{MERGER}}_t = \gamma_0 + \gamma_1 R_{\text{SP500}}_t + u_t \]
\[ u_t \sim N(0, h_t) \]
\[ h_t = \alpha_0 + \alpha_1 u^2_{t-1} + \beta_1 \sigma^2_{t-1} + \lambda u^2_{t-1} dt_{t-1}. \]

Interpret the parameter estimates and sketch the news impact curve. Test the significance of the threshold parameter \(\lambda\) and interpret the result.

(c) Estimate the following CAPM with time-varying variance for the Merger hedge fund

\[ R_{\text{MERGER}}_t = \gamma_0 + \gamma_1 R_{\text{SP500}}_t + \theta h_t + u_t \]
\[ u_t \sim N(0, h_t) \]
\[ h_t = \alpha_0 + \alpha_1 u^2_{t-1} + \beta_1 \sigma^2_{t-1} + \lambda u^2_{t-1} dt_{t-1}. \]

Conduct a test of time variation in the risk on the hedge fund.

(d) Repeat parts (a) to (c) for the other six hedge funds. How successful are the hedge funds in minimizing exposure from systematic risk to the market? Discuss.

8. Estimating Foreign Exchange Market Volatility

This question is based on the file \texttt{HOUR.WF1} which contains hourly data on the British pound (BP) and the Deutchmark (DM) exchange rates, relative to the US, over the period 0.00am January 1st, 1986 to 11.00am July 15th 1986.

(a) For each exchange rate \(E\), compute the returns as

\[ R_E = \log(E) - \log(E(-1)) \]

(b) Test for an ARCH(1) effect in foreign exchange returns.

(c) Test for an ARCH(2) effect in foreign exchange returns.

(d) Provided that there is a significant ARCH effect estimate a \textit{GARCH}(1,1) model.
11.10. EXERCISES

(e) Provided that there is a significant ARCH effect estimate a GARCH($1, 1$) – $M$ model where the conditional mean is a function of the conditional standard deviation.

(f) Provided that there is a significant ARCH effect estimate a GARCH($1, 1$) – $M$ model where the conditional mean is a function of the conditional variance.

(g) Discuss the volatility of foreign exchange returns by comparing the estimated GARCH standard deviation from the various models with actual returns.

9. Minimum Variance Portfolio Model with Time-Varying Weights

This question is based on the EViews file HEDGE.WF1 used in Question 1. Let $R_1$ and $R_2$ be respectively the percentage continuously compounded daily returns on the Convertible Arbitrage hedge fund and the Distressed hedge fund

\[
R_1 = R_{\text{CONVERTIBLE}} \quad R_2 = R_{\text{DISTRESSED}}
\]

(a) Calculate the variance-covariance matrix of $R_1$ and $R_2$ and interpret the elements.

(b) Consider a portfolio containing shares on the Convertible Arbitrage hedge fund and the Distressed hedge fund with respective weights $w_1$ and $w_2 = 1 - w_1$. The variance of the portfolio is given by

\[
\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1w_2 \sigma_{1,2},
\]

where $\sigma_1^2$ and $\sigma_2^2$ are the respective Convertible Arbitrage hedge fund and the Distressed hedge fund return variances, and $\sigma_{1,2}$ is the covariance between the returns on the two funds. The minimum variance portfolio is given by setting the weight on the Convertible Arbitrage hedge fund in the portfolio to

\[
w_1 = \frac{\sigma_2^2 - \sigma_{1,2}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}}.
\]

Using the estimates of the variances and the covariance obtained in part (a), estimate $w_1$ and hence $w_2$. Interpret the results.

(c) Estimate the following regression equation

\[
R_{2t} = \beta_0 + \beta_1 (R_{2t} - R_{1t}) + u_t,
\]

where $u_t$ is an error term and $\beta_0$ and $\beta_1$ are parameters. Hence identify the relationship between the estimate of $\hat{\beta}_1$ obtained in this question and the estimate of $w_1$ obtained in part (b). Using this information, test the statistical significance of minimizing risk through portfolio diversification.
Consider the minimum variance portfolio model with time-varying weights

\[ R_{2t} = \beta_0 + \beta_{1,t-1} (R_{2t} - R_{1t}) + u_t, \]

where

\[ \beta_{1,t-1} = \frac{\text{Cov}(R_{2t}, R_{2t} - R_{1t} | \Omega_{t-1})}{\text{Var}(R_{2t} - R_{1t} | \Omega_{t-1})}, \]

and \( \Omega_{t-1} \) is the information set at time \( t - 1 \).

Compute the following

\[ E1 = R1 - @\text{MEAN} (R1) \]
\[ E2 = R2 - @\text{MEAN} (R2) \]

What is the relationship between the sample mean of \( R1 \) and \( E1 \)?

What is the relationship between the sample variance of \( R1 \) and \( E1 \)?

Estimate the following regression equations

\[ (E_{2t} - E_{1t})^2 = \alpha_0 + \alpha_1 (E_{2t-1} - E_{1t-1})^2 + v_{1,t}, \]
\[ (E_{2t}) (E_{2t} - E_{1t}) = \beta_0 + \beta_1 (E_{2t-1}) (E_{2t-1} - E_{1t-1}) + v_{2,t}, \]

and use these results to estimate the time-varying weight \( \beta_{1,t-1} \).

Compare the estimate of \( \beta_{1,t-1} \) with the constant weight estimate obtained in part (c).
Investing in financial securities is risky because of the variability of future returns. As seen in Chapter 11, the time-varying variance of financial asset returns is relatively easy to model for the case of an individual asset. Understanding the co-movements of financial returns is also of great practical importance. For example, asset pricing depends on the covariance of the assets in a portfolio and risk management and asset allocation relate for instance to finding and updating optimal hedging positions. In addition, forecasts of conditional covariance matrices of financial returns continue to influence the management of large amounts of funds under management. A 2012 survey of 139 North American investment managers representing $12 trillion worth of assets under management reported that the majority of managers use volatility and correlation forecasts to construct equity portfolios (Amenc, Goltz, Tang and Vaidyanathan, 2012). It follows therefore that specifying, estimating and forecasting using multivariate models of the conditional covariance matrix of a portfolio of assets are of great practical importance.

Let \( r_t \) denote a vector of time series \( r_t = \{r_{1t}, \cdots, r_{Nt}\} \) and let \( I_{t-1} \) represent the information set available at time \( t-1 \), the fundamental multivariate time-varying conditional variance problem addressed in this chapter is

\[
\begin{align*}
    r_t &= \mathbb{E}(r_t | I_{t-1}) + u_t \\
    \text{var}(u_t | I_{t-1}) &= H_t.
\end{align*}
\]  

(12.1)

The focus here will be on the specification of \( H_t \) and, as a consequence, the mean equation will largely be ignored by invoking that simplifying assumption that \( r_t \) has been recentered to have zero mean. The natural approach to modelling a time-varying conditional covariance matrix is to extend the univariate GARCH framework to a multivariate version in which both variances and covariances are allowed to be time-varying. In addition to the the multivariate GARCH (MGARCH) models more recent developments focussing on
the dynamics of correlations between asset will be also be discussed. The central feature relating to volatility from Chapter 11, namely that it is regarded as unobservable, is maintained in these multivariate extensions. The relatively new areas of research in which realised volatility and realised covariance, which serve observable measures for volatility and covariance, are proposed are dealt with in Chapter ???. The problem of forecasting multivariate volatility will be postponed until these new observable proxies for volatility have been introduced.

12.1 Motivation

12.1.1 Time-Varying Beta Risk

In Chapter 3 the beta risk of asset $i$, is defined as

$$
\beta = \frac{\text{cov}(r_{it} - r_{ft}, r_{mt} - r_{ft})}{\text{var}(r_{mt} - r_{ft})} = \frac{\mathbb{E}[(r_{it} - r_{ft})(r_{mt} - r_{ft})]}{\mathbb{E}[(r_{mt} - r_{ft})^2]}
$$

where $r_{it} - r_{ft}$ is the excess return on the asset relative to the risk-free rate given by $r_{ft}$ and $r_{mt} - r_{ft}$ is the corresponding excess return on the market portfolio. Using the monthly data set on United States stocks for the period April 1990 to July 2004 ($T = 172$) introduced in Chapter 3, the constant beta risk for the stock Microsoft is easily estimated using the CAPM least squares regression. The estimate of the constant beta risk is $\hat{\beta} = 1.447$.

The key restriction of constant beta risk may, however, be unrealistic. For example, the early 2000s was the period of the DotCom bubble and the beta risk of a technology stock like Microsoft could have been affected. Consequently, it would be desirable to be able to relax this restriction and allow beta to be time-varying. The specification of beta then becomes

$$
\beta_t = \frac{\mathbb{E}_{t-1}[(r_{it} - r_{ft})(r_{mt} - r_{ft})]}{\mathbb{E}_{t-1}[(r_{mt} - r_{ft})^2]}
$$

(12.2)

where $\mathbb{E}_{t-1}[(r_{it} - r_{ft})(r_{mt} - r_{ft})]$ is the conditional covariance between the excess returns of Microsoft and the market based on information at time $t - 1$ and $\mathbb{E}_{t-1}[(r_{mt} - r_{ft})^2]$ is the conditional variance of the market excess return. The problem is how to operationalise the estimate of $\beta_t$ in equation (12.2).

From Chapter 11 the conditional covariance of Microsoft returns can be estimated using a simple GARCH model. Similarly, the same approach can be used to estimate the conditional covariance of the returns to the market portfolio, in this case proxied by the returns to the S&P 500 index. The problem is the conditional covariance term $\mathbb{E}_{t-1}[(r_{it} - r_{ft})(r_{mt} - r_{ft})]$. Clearly what is required is a bivariate model in which conditional variances of the constituent assets and their conditional covariance are all modelled simultaneously.

One approach would be to adopt the the multivariate versions of the historical variance and the exponentially weighted moving average estimate
discussed in Chapter 11. It the excess returns are collected into the vector
\( r_t = [r_{1t}, r_{mt}]' \) then these measures are given as follows.

(i) **Historical Variance:**

The multivariate version of the historical estimate of the conditional
covariance matrix of \( r_t \) is

\[
H_t = \frac{1}{M} \sum_{j=1}^{T} r_{t-j}'r_{t-j}.
\]

The unconditional covariance matrix results if \( M = T \). The forecast of
volatility for \( k \)-periods ahead is simply given by the current value \( H_t \),
irrespective of the value of \( k \).

(ii) **Exponential Weighted Moving Average:**

The multivariate version of the exponentially weighted moving average
estimate of conditional covariance is given by

\[
H_t = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j r_{t-j}'r_{t-j} = (1 - \lambda)T_t + \lambda H_t
\]

where, as before, \( \lambda \) is the delay parameter.

These two estimates suffer from exactly the same shortcomings as their uni-
variate counterparts discussed in Chapter 11 and will not be pursued here.

The search for a viable approach to the estimation of time-varying conditional
covariance matrices starts by estimating simple GARCH(1,1) models for the
excess returns to Microsoft, \( r_{1t} - r_{ft} \), and the excess returns to the S&P 500
index, \( r_{mt} - r_{ft} \). This exercise yields the following estimated equations for the
conditional variances of Microsoft, \( h_{1t} \), and the S&P 500, \( h_{mt} \), respectively,

\[
\begin{align*}
  h_{1t} &= 5.955 + 0.141 \hat{u}_{1t}^2 + 0.802 h_{t-1} \\
  h_{mt} &= 0.573 + 0.112 \hat{u}_{2t}^2 + 0.858 h_{mt-1},
\end{align*}
\]

in which \( \hat{u}_{1t} \) and \( \hat{u}_{2t} \) are the residuals from the respective mean equations
which contain only a constant term. The conditional variance equations de-
scribe time variation in the variance of the individual assets but do not cap-
ture time variation in the conditional covariance between the two assets, \( h_{1mt} \).

From Chapter 2, recall that this the correlation between these two random
variables is given by

\[
\rho = \frac{h_{1m}}{\sqrt{h_{1t}} \sqrt{h_{mt}}}
\]

where \( h_{1m} \) is the conditional covariance between them. This unconditional
measure can be made into a time-varying one by allowing the variances and
covariances to vary so that

\[
\rho_t = \frac{h_{1mt}}{\sqrt{h_{mt}} \sqrt{h_{1t}}}
\]
If the additional restriction of a constant correlation is imposed, \( \rho_t = \bar{\rho} \) then an estimate of the time-varying conditional covariance is given by

\[
h_{1mt} = \bar{\rho} \sqrt{h_{mt}} \sqrt{h_{1t}}.
\]

A reasonable choice for \( \bar{\rho} \) is the sample correlation between the excess returns to Microsoft and the excess returns on the market. Using the estimated equations for \( h_{mt} \) and \( h_{1t} \) and \( \bar{\rho} = 0.5804 \), the sample correlation coefficient, a series for the conditional covariance, \( h_{1mt} \), can be computed. The two conditional variances and the conditional covariance computed in this way are shown in Figure 12.1. Although this a very simple approach, the major insight it provides has proved important in developing workable multivariate GARCH models which are discussed in Section 12.4.

![Figure 12.1: Conditional variances (top panel) and covariance (bottom panel) of Microsoft and the S&P500 index. The data are monthly for the period April 1990 to July 2004 (T = 172).](image)

It is apparent that the conditional variances change over the sample with Microsoft showing a marked increase in volatility at the time of the DotCom bubble in the early 2000s. Figure 12.1 also shows that the covariance and the variance of the market tend to decrease in-step with each other in the first half of the sample, but appear to be out of alignment in the second half of the sample.

Using these estimates for the conditional covariance, \( h_{1mt} \), and the conditional variance of the market, \( h_{mt} \), an estimate of time-varying beta risk is plotted in Figure 12.2 and superimposed on the constant estimate of beta risk. There are
12.1. MOTIVATION

Figure 12.2: Estimate of time-varying beta risk for Microsoft based on the assumption of a constant correlation between Microsoft and the S&P 500 index. The constant beta risk estimated from a CAPM model of 1.447 is shown as the dashed line. The data are monthly for the period April 1990 to July 2004 ($T = 172$).

some very large changes in the beta risk of Microsoft, ranging from around 0.2 at the end of 1995 to nearly 2.5 at the time of the DotCom bubble in the middle of 2000. The sample average is 1.196 which is a little lower than the constant estimate of beta risk given by 1.447.

12.1.2 Time-varying Portfolio Weights

Another simple ($N = 2$) but important example where knowledge of the entire time-varying conditional covariance matrix is required is the construction of the optimal portfolio. The minimum variance portfolio containing two assets with returns $r_{1t}$ and $r_{2t}$ has optimal weights given by

$$w_1 = \frac{\text{var} (r_{2t}) - \text{cov} (r_{1t}, r_{2t})}{\text{var} (r_{1t}) + \text{var} (r_{2t}) - 2 \text{cov} (r_{1t}, r_{2t})}$$

$$w_2 = \frac{\text{var} (r_{1t}) - \text{cov} (r_{1t}, r_{2t})}{\text{var} (r_{1t}) + \text{var} (r_{2t}) - 2 \text{cov} (r_{1t}, r_{2t})}$$

where $w_1$ is the optimal weight allocated to asset 1 in the portfolio and $w_2$ is the corresponding weight on asset 2.

Even in this simple case of two assets, $N = 2$, construction of a model based estimate of the covariance matrix is not entirely straightforward. There are two time-varying variances and one time-varying covariance

$$\text{var} (r_{1t}) : h_{11t} = E_t - [E_t - E_{t-1} [r_{1t}]]^2$$

$$\text{var} (r_{2t}) : h_{22t} = E_t - [E_t - E_{t-1} [r_{2t}]]^2$$

$$\text{cov} (r_{12}, r_{2t}) : h_{12t} = E_t - [E_t - E_{t-1} [r_{1t}]] (r_{2t} - E_{t-1} [r_{2t}])$$,
so that the conditional covariance matrix is

\[
H_t = \begin{bmatrix}
h_{11t} & h_{12t} \\
h_{21t} & h_{22t}
\end{bmatrix} = \begin{bmatrix}
h_{11t} & h_{12t} \\
h_{12t} & h_{22t}
\end{bmatrix}
\] (12.3)

because of symmetry.

Figure 12.3: Estimates of time-varying weights for portfolio comprising Microsoft and Walmart. The weights are constructed based on the assumption of a constant correlation between Microsoft and Walmart. The data are monthly for the period April 1990 to July 2004 (\(T = 172\)).

The time-varying portfolio weights, together with the optimal constant weights are shown in Figure 12.3. Obviously as the weights in this two asset portfolio sum to 1, the time varying weights are mirror images of each other. The one important feature of these weights is that the constant values obtained from the regression approach in Chapter 3 are completely dominated by the DotCom crisis. The time-varying versions show that Microsoft should have received the higher weight in the portfolio for the entire 10-year period leading up to 2000. This result demonstrates the usefulness of modelling time-variation in the variances and covariances of financial assets explicitly rather than relying on the simplifying assumption of constant relationships.

### 12.2 Heatwaves and Meteor Showers

An early empirical application aimed at extending the GARCH framework into multi-dimensions, while still using the tools developed in the univariate context, examined how volatility is transmitted through different regions of the world during the course of a global financial trading day (Ito, 1987; Ito and Roley, 1987; Engle, Ito and Lin (1990). The approach is to partition each 24 hour period (calendar day) into a three major trading zones, namely, Japan
12.2. HEATWAVES AND METEOR SHOWERS

(12am to 7am GMT), Europe (7am to 12:30pm GMT) and the United States (12:30pm to 9pm GMT), which may be illustrated as follows:

<table>
<thead>
<tr>
<th></th>
<th>Japan</th>
<th>Europe</th>
<th>U.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>12am-7am</td>
<td>12am</td>
<td>7am</td>
<td>12pm</td>
</tr>
<tr>
<td>7am-12pm</td>
<td></td>
<td>12pm</td>
<td>9pm</td>
</tr>
</tbody>
</table>

One Trading Day

Note that there are other ways of carving up the global trading day, see for example Dungey, Fakhrutdinova and Goodhart (2009), but the main thrust of the argument remains the same irrespective of the minor adjustments to this definition.

The calendar structure implied by the global trading day defines a number of restrictions on a three equation system which uses a simple GARCH(1,1) model for modelling the conditional variance in each of the trading zones. Define $r_{1t}$, $r_{2t}$ and $r_{3t}$ as the daily returns to the Japanese zone, the European zone and the United States zone, respectively. The model is

$$
\begin{bmatrix}
  r_{1t} \\
  r_{2t} \\
  r_{3t}
\end{bmatrix} =
\begin{bmatrix}
  u_{1t} \\
  u_{2t} \\
  u_{3t}
\end{bmatrix},
\begin{bmatrix}
  u_{1t} \\
  u_{2t} \\
  u_{3t}
\end{bmatrix} \sim N\left(\begin{bmatrix}
  0 \\
  0 \\
  0
\end{bmatrix},
\begin{bmatrix}
  h_{1t} & 0 & 0 \\
  0 & h_{2t} & 0 \\
  0 & 0 & h_{3t}
\end{bmatrix}\right),
$$

$$
\begin{bmatrix}
  h_{1t} \\
  h_{2t} \\
  h_{3t}
\end{bmatrix} =
\begin{bmatrix}
  \alpha_{10} \\
  \alpha_{20} \\
  \alpha_{30}
\end{bmatrix} +
\begin{bmatrix}
  0 & 0 & 0 \\
  \alpha_{21} & 0 & 0 \\
  \alpha_{31} & \alpha_{32} & 0
\end{bmatrix}
\begin{bmatrix}
  u_{1t}^2 \\
  u_{2t}^2 \\
  u_{3t}^2
\end{bmatrix} +
\begin{bmatrix}
  \beta_{11} & 0 & 0 \\
  0 & \beta_{22} & 0 \\
  0 & 0 & \beta_{33}
\end{bmatrix}
\begin{bmatrix}
  h_{1t-1} \\
  h_{2t-1} \\
  h_{3t-1}
\end{bmatrix} +
\begin{bmatrix}
  \gamma_{11} & \gamma_{12} & \gamma_{13} \\
  0 & \gamma_{22} & \gamma_{23} \\
  0 & 0 & \gamma_{33}
\end{bmatrix}
\begin{bmatrix}
  u_{1t-1}^2 \\
  u_{2t-1}^2 \\
  u_{3t-1}^2
\end{bmatrix}.
$$

(12.4)

The calendar structure of the global trading day is now apparent. New developments at the start of the global trading day in Japan, $u_{1t}$, can potentially influence volatility in Europe and the United States via the coefficients $\alpha_{21}$ and $\alpha_{31}$. Similarly news from Europe, $u_{2t}$, can influence volatility in the United States on the same global trading day, $\alpha_{32}$. The natural calendar structure, however, implies that events in the United States will be transmitted to Japan only on the following day. The restrictions on the $\gamma_{ij}$ on the lagged innovations, which require the matrix to be upper diagonal, imply that all information originates during United States trading times.

While this model looks very much like a multivariate GARCH model, there is no contemporaneous conditional covariance because the regions are defined to be non-overlapping. For this reason single equation estimation of the model by the maximum likelihood can be performed on each zone using the estimation methods outlined in Chapter 11. The aim is to examine international linkages in volatility between these regions and investigate in partic-
ular two patterns as possible descriptors of international volatility transmission.

(i) **Heatwave**
Volatility in any one region is primarily a function of the previous day’s volatility in the same region.

(ii) **Meteor Shower**
Volatility in one region is driven primarily by volatility in the region immediately preceding it in terms of calendar time.

To test this model continuously-traded, high-frequency data is used to construct returns to the Euro-Dollar exchange rate futures contracts traded on the Chicago Mercantile Exchange for period 3 January 2005 to 28 February 2013 ($T = 1976$ trading days). The return in each zone is calculated as the difference between the last and the first transaction price for the time period in which the zone trades, normalised by the number of hours for which the zone trades, given by

$$r_{it} = \frac{\log P_{ct} - \log P_{ot}}{\sqrt{nh_i}}, \quad i = 1, 2, 3 \tag{12.5}$$

in which, $P_{ct}$ is a closing price of the futures contract in zone $i$ on day $t$, and $P_{ot}$ is the opening price of the contract in zone $i$ on day $t$ and $nh_i$ is the number of hours for which zone $i$ trades. Descriptive statistics for the returns from each zone are presented in Table 12.1. While none of the returns series from the three zones exhibit large degrees of skewness, they all exhibit excess kurtosis. Formal testing reveals that all the series strong ARCH effects at the 5% level.

Table 12.1


<table>
<thead>
<tr>
<th></th>
<th>Japan</th>
<th>Europe</th>
<th>U.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.038</td>
<td>-0.086</td>
<td>0.056</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>1.214</td>
<td>1.873</td>
<td>1.876</td>
</tr>
<tr>
<td>Max.</td>
<td>6.707</td>
<td>8.147</td>
<td>11.547</td>
</tr>
<tr>
<td>Min.</td>
<td>-5.821</td>
<td>-12.754</td>
<td>-9.307</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.014</td>
<td>-0.293</td>
<td>0.049</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.640</td>
<td>5.874</td>
<td>5.378</td>
</tr>
</tbody>
</table>

The estimation results for equation (12.4) based on the foreign exchange market futures data are reported in Table 12.2. Note that the constant term in
the variance equation is suppressed. There are two general conclusions that emerge from inspection of these results.

1. All the lagged conditional variance terms, $h_{t-1}^i$, are statistically significant and of the order of 0.9, an estimate which is commonly obtained in univariate GARCH models applied to financial returns. The statistical significance of these terms is consistent with the heatwave hypothesis being part of the explanation of the patterns in global volatility.

2. The meteor shower effect (diurnal effect of the news) is also important: Japanese news affects Europe and European news affects the United States on the same trading day. Note that in the case of Japan, the meteor shower effect shows up in the significance of the lagged influence of the United States innovations, $u_{3t-1}^2$, on Japan.

It seems clear, therefore, that the pattern of volatility interaction in global foreign markets is a combination of both heat waves and meteor showers. There is no support for the conclusion that either one of these patterns dominates.

Table 12.2

<table>
<thead>
<tr>
<th></th>
<th>Japan</th>
<th>Europe</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_{1t}^2$</td>
<td>-</td>
<td>0.0356*</td>
<td>-0.0320</td>
</tr>
<tr>
<td>$u_{2t}^2$</td>
<td>-</td>
<td>-</td>
<td>0.0646*</td>
</tr>
<tr>
<td>$u_{3t}^2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$u_{1t-1}^2$</td>
<td>0.0850*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$u_{2t-1}^2$</td>
<td>-0.0258</td>
<td>0.0850*</td>
<td>-</td>
</tr>
<tr>
<td>$u_{3t-1}^2$</td>
<td>0.0764*</td>
<td>-0.0027*</td>
<td>0.0814*</td>
</tr>
<tr>
<td>$h_{t-1}$</td>
<td>0.8832*</td>
<td>0.9024*</td>
<td>0.8992*</td>
</tr>
</tbody>
</table>

This pattern of interaction suggests the transmission of news between different regions of the world on the same trading day is a potentially important explanation of volatility. This result simply adds to the compelling argument to support the importance of being able to estimate the full covariance matrix of a system of $M$ financial asset returns, $r_t = \{r_{1t}, \ldots, r_{Mt}\}$, given that continuous trading on many important financial exchanges now makes the artefact of dividing the global day into different trading zones largely irrelevant.
12.3 Multivariate Conditional Covariance

Consider a set of \( N \) asset returns, \( r_t = \{ r_{1t}, \cdots, r_{Nt} \} \), where without loss of generality the returns are assumed to be centered to have zero mean. A time-varying estimate of the conditional covariance matrix requires construction of the matrix

\[
H_t = \begin{bmatrix}
  h_{11t} & h_{12t} & \cdots & h_{1Nt} \\
  h_{21t} & h_{22t} & \cdots & h_{2Nt} \\
  \vdots & \vdots & \ddots & \vdots \\
  h_{N1t} & h_{N2t} & \cdots & h_{NNt}
\end{bmatrix},
\]

which is a symmetric matrix because \( h_{ijt} = h_{jit} \) so that there are \( N(N+1)/2 \) unique elements. A good estimate of \( H_t \) satisfies two conditions.

(i) **Positive Definiteness**

All conditional variances and covariances must be positive for all \( t \). This restriction requires that the conditional covariance matrix \( H_t \) be positive definite. For \( N = 2 \) in equation, the conditions for positive definiteness are

\[
\begin{align*}
  h_{11t} &> 0 \\
  h_{11t}h_{22t} - h_{12t}^2 &> 0
\end{align*}
\]

The second condition may be rewritten as

\[
-1 < \frac{h_{12t}}{\sqrt{h_{11t}h_{22t}}} < 1
\]

so the correlation needs to be between \(-1\) and 1 at every \( t \). Ensuring that this condition is met is not straightforward, particularly as the dimension grows.

(ii) **Parameter Dimension**

Consider a multivariate version of the simple AR(1) model of squared returns introduced in Chapter 11 to motivate the ARCH model. In theory there is no reason why conditional variances and covariances should not be functions of shocks from all other assets, an observation that suggests the following specification

\[
\begin{align*}
r_{1t}^2 &= a_0 + a_1 r_{1t-1}^2 + a_2 r_{2t-1}^2 + a_3 r_{1t-1}r_{2t-1} + v_{1t} \\
r_{2t}^2 &= \beta_0 + \beta_1 r_{1t-1}^2 + \beta_2 r_{2t-1}^2 + \beta_3 r_{1t-1}r_{2t-1} + v_{2t} \\
r_{1t}r_{2t} &= \gamma_0 + \gamma_1 r_{1t-1}^2 + \gamma_2 r_{2t-1}^2 + \gamma_3 r_{1t-1}r_{2t-1} + v_{3t}
\end{align*}
\]

This simple specification for two returns \( r_{1t} \) and \( r_{2t} \) already has 12 parameters to estimate. For large portfolios, say \( N = 50 \) or 100, the model becomes very difficult to estimate as there are potentially too many parameters.
12.3. MULTIVARIATE CONDITIONAL COVARIANCE

Dealing with these two problems has seen the development of a number of multivariate GARCH specifications which are designed to provide a specification that is flexible enough to model the dynamics of volatility and co-volatility over time, while ensuring both that the covariance matrix is positive definite and controlling the dimension of the parameter space.

12.3.1 The VECH Model

This is the first MGARCH specification proposed in the literature. The specification is based on using just the $N(N + 1)/2$ unique elements of the covariance matrix $H_t$, which this is achieved by using the vech ($\cdot$) matrix operator. For $N = 2$, the vech operator is defined as follows:

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{bmatrix} \implies \text{vech}(H_t) = \begin{bmatrix} h_{11t} \\ h_{12t} \\ h_{22t} \end{bmatrix}$$

For $N = 3$ the operation of the vech operator yields

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} & h_{13t} \\ h_{12t} & h_{22t} & h_{23t} \\ h_{13t} & h_{23t} & h_{33t} \end{bmatrix} \implies \text{vech}(H_t) = \begin{bmatrix} h_{11t} \\ h_{12t} \\ h_{13t} \\ h_{22t} \\ h_{23t} \\ h_{33t} \end{bmatrix}.$$

The equation governing the dynamics of the conditional variance for an MARCH(1,1) model using the VECH specification is

$$\text{vech}(H_t) = C + A \text{vech}(u_{t-1}u_{t-1}') + B \text{vech}(H_{t-1})$$

which, in the case of $N = 2$ becomes

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

after application of the vech(\cdot) operator.

The VECH model represents the first attempt at building an MGARCH model but it has two important drawbacks. The first is that $H_t$ is not necessarily positive definite, even if $c_i > 0$ for $i = 1, 2, 3$. The second is that the number of parameters to be estimated increases exponentially as the dimension increases.

For example, in the case of $N = 2$ there are

$$\frac{N(N + 1)}{2} + 2\left(\frac{N(N + 1)}{2}\right)^2 = \frac{2(2 + 1)}{2} + 2\left(\frac{2(2 + 1)}{2}\right)^2 = 21,$$
variance parameters to be estimated. If \( N = 3 \) it is easily verified that there are 78 variance parameters to be estimated. Consequently the VEC version of the MGARCH model is not used much in practice, although a diagonal version, in which all the off-diagonal elements of the matrices \( C, A \) and \( B \) are zero, has proved more popular.

12.3.2 The BEKK Model

Letting \( H_t \) be the conditional covariance matrix at time \( t \), and \( u_t \) the vector of disturbances, the BEKK specification for a MGARCH(1,1) model is

\[
H_t = C C' + A u_{t-1} u_{t-1}' A' + B H_{t-1} B'
\]

where \( C \) is a \((N \times N)\) lower triangular matrix of unknown parameters, and \( A \) and \( B \) are \((N \times N)\) matrices each containing \( N^2 \) unknown parameters associated with the lagged disturbances and the lagged conditional covariance matrix respectively.

There are three different flavours of BEKK model, each of which is now spelled out for the case \( N = 2 \).

(i) **Asymmetric BEKK Model**

The parameter matrices are

\[
C = \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}
\]

This is the asymmetric BEKK model as \( a_{12} \neq a_{21}, b_{12} \neq b_{21} \).

(ii) **Symmetric BEKK Model**

The restrictions \( a_{12} = a_{21} \) and \( b_{12} = b_{21} \), are imposed so

\[
C = \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{bmatrix}
\]

(iii) **Diagonal BEKK Model**

The restrictions \( a_{12} = a_{21} = 0 \) and \( b_{12} = b_{21} = 0 \), are imposed so

\[
C = \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}
\]

A special case of the BEKK model is where there is just the one variable \((N = 1)\), so the parameter matrices become scalars

\[
C = [c_{11}], \quad A = [a_{11}], \quad B = [b_{11}].
\]

The conditional covariance matrix now reduces to a scalar given by

\[
H_t = h_{11t} = c_{11}^2 + a_{11}^2 u_{t-1}^2 + b_{11}^2 h_{11t-1}
\]
which is simply the univariate GARCH(1,1) model discussed earlier with the difference that the parameters are squared. This last feature of the BEKK model highlights the motivation behind the choice of the specification. In the univariate case the conditional variance is constrained to be positive because all terms on the RHS are positive, that is,

$$u_{t-1}^2 > 0, \quad \{c_{11}^2, a_{11}^2, b_{11}^2\} > 0, \quad h_{11t-1} > 0.$$  

The BEKK specification has the advantage that it solves the first problem of positive definiteness, but not necessarily the second problem when the dimension of the model is relatively large. For most empirical work using this model is based on $N$ less than 10 and often it is $N = 2$ or $N = 3$. In addition, the unrestricted model contains parameters that do not represent directly the impact of $u_{t-1}$ or $H_{t-1}$ on the elements of $H_t$ and this makes it hard to interpret the parameters of a BEKK model.

### 12.4 Multivariate Conditional Correlation

In Section 12.1 the simple assumption of constant correlation between two financial assets was made in order to develop an estimate of the conditional covariance between them. This assumption has proved to be a good starting point for the development of multivariate models of conditional correlation which have become the preferred way in which to model the conditional covariance between systems of financial assets.

#### 12.4.1 Constant Conditional Correlation (CCC) Model

As established in Chapter 2, a covariance is equal to correlation times the respective standard deviations. The conditional covariance matrix is therefore defined as

$$H_t = S_t R S_t,$$

where $S_t$ is a $(N \times N)$ diagonal matrix of time-varying standard deviations and $R$ is a $(N \times N)$ matrix of constant correlations. This generalises the simple assumption made in Section 12.1 from the bivariate to the multivariate case. The standard deviation matrix $S_t$ is a diagonal matrix with the time-varying standard deviations down the main diagonal

$$S_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 & \cdots \\ \vdots & \ddots & \ddots \\ 0 & \cdots & \sqrt{h_{nt}} \end{bmatrix},$$

which are the square roots of univariate GARCH specifications

$$h_{it} = \alpha_{0i} + \alpha_{1i}u_{i,t-1}^2 + \beta_{1i}h_{it-1}, \quad i = 1, 2, \ldots, N.$$
The constant correlation matrix $R$ is defined as
\[
R = \begin{bmatrix}
1 & r_{12} & \cdots & r_{1N} \\
r_{21} & 1 & \cdots & r_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
r_{N1} & r_{N2} & \cdots & 1
\end{bmatrix} = \text{diag}(Q)^{-1/2} \overline{Q} \text{diag}(Q)^{-1/2},
\]
which is a symmetric matrix. The matrix $Q$ is defined as
\[
Q = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix}
z_{1t}^2 & z_{1t}z_{2t} & \cdots & z_{1t}z_{Nt} \\
z_{2t}z_{1t} & z_{2t}^2 & \cdots & z_{2t}z_{Nt} \\
\vdots & \vdots & \ddots & \vdots \\
z_{Nt}z_{1t} & z_{Nt}z_{2t} & \cdots & z_{Nt}^2
\end{bmatrix}
\]
(12.6)
and $z_{it}$ represents the standardised disturbances from the GARCH models
\[
z_{it} = \frac{u_{it}}{\sqrt{h_{it}}}
\]
For illustrative purposes, it is useful to outline the construction of the conditional covariance matrix of the constant model for the 2 asset case. The usual notation to denote estimates will be suppressed for simplicity. Use the stationary univariate GARCH(1, 1) models for the two assets
\[
h_{1t} = \alpha_0 + \alpha_{11} u_{1,t-1}^2 + \beta_{11} h_{1,t-1} \\
h_{2t} = \alpha_2 + \alpha_{12} u_{2,t-1}^2 + \beta_{12} h_{2,t-1},
\]
construct the standard deviation matrix
\[
S_t = \begin{bmatrix}
\sqrt{h_{1t}} & 0 \\
0 & \sqrt{h_{2t}}
\end{bmatrix}.
\]
Compute the standardised disturbances
\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} = S_t^{-1} u_t = \begin{bmatrix}
\frac{1}{\sqrt{h_{1t}}} & 0 \\
0 & \frac{1}{\sqrt{h_{2t}}}
\end{bmatrix} \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix} = \begin{bmatrix}
\frac{u_{1t}}{\sqrt{h_{1t}}} \\
\frac{u_{2t}}{\sqrt{h_{2t}}}
\end{bmatrix}
\]
and $\overline{Q}$
\[
\overline{Q} = \begin{bmatrix}
\overline{q}_{11} & \overline{q}_{12} \\
\overline{q}_{12} & \overline{q}_{22}
\end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix}
z_{1t}^2 & z_{1t}z_{2t} \\
z_{2t}z_{1t} & z_{2t}^2
\end{bmatrix}.
\]
Use $\overline{Q}$ to estimate the constant correlation matrix as
\[
R = \begin{bmatrix}
1 & \frac{\overline{q}_{12}}{\sqrt{\overline{q}_{11}\overline{q}_{22}}} \\
\frac{\overline{q}_{12}}{\sqrt{\overline{q}_{11}\overline{q}_{22}}} & 1
\end{bmatrix}.
\]
12.4. MULTIVARIATE CONDITIONAL CORRELATION

Finally, construct the conditional covariance matrix as

\[
H_t = \begin{bmatrix}
\sqrt{h_{1t}} & 0 \\
0 & \sqrt{h_{2t}}
\end{bmatrix}
\begin{bmatrix}
1 & \frac{\tilde{q}_{12}}{\sqrt{q_{11}q_{22}}} \\
\frac{\tilde{q}_{12}}{\sqrt{q_{11}q_{22}}} & 1
\end{bmatrix}
\begin{bmatrix}
\sqrt{h_{1t}} & 0 \\
0 & \sqrt{h_{2t}}
\end{bmatrix}
\]

The CCC model solves both the positive definiteness and the dimensionality issues. However, the assumption of correlations being constant over the sample is potentially restrictive. Consequently subsequent developments have attempted to relax this assumption.

12.4.2 Dynamic Conditional Correlation (DCC) Models

The Dynamic Conditional Correlation (DCC) model relaxes the assumption that the correlations are constant over time and introduces some simple dynamics to the time evolution of the matrix. The conditional covariance matrix is now specified as

\[
H_t = S_t R_t S_t
\]

where \( R_t \) is now a \((N \times N)\) conditional correlation matrix. As before \( S_t \) is a diagonal matrix containing the conditional standard deviations

\[
S_t = \begin{bmatrix}
\sqrt{h_{1t}} & 0 \\
0 & \cdots \\
0 & \sqrt{h_{Nt}}
\end{bmatrix}
\]

where the conditional variances have univariate GARCH representations

\[
h_{it} = \alpha_0 + \alpha_1 u_{i,t-1}^2 + \beta h_{i,t-1}, \quad i = 1, 2, \ldots, N.
\]

The conditional correlation matrix \( R_t \) is given by

\[
R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2},
\]

where \( Q_t \) has a GARCH(1,1) type specification

\[
Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha z_{i,t-1} z_{i,t-1}^t + \beta Q_{t-1}.
\]

The definition of the intercept in the dynamics of \( Q_t \) as \((1 - \alpha - \beta) \bar{Q}\) corresponds to the idea of variance targeting introduced in Engle and Mezrich (1996), in which specifying the intercept in this way reduces the difficulty of
the estimation problem because the intercept no longer has to be estimated separately. In fact, this specification of the dynamics of the conditional correlation matrix is particularly parsimonious because it contains just two unknown scalar parameters, \( \alpha \) and \( \beta \) (with \( \alpha + \beta < 1 \)), the standardised residuals

\[ z_{it} = \frac{u_{it}}{\sqrt{h_{it}}} \]

and the quasi-correlation matrix (Aielli, 2009), \( Q \), given by

\[ Q = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} z_{1t}^2 & z_{1t}z_{2t} & \cdots & z_{1t}z_{Nt} \\ z_{2t}z_{1t} & z_{2t}^2 & \cdots & z_{2t}z_{Nt} \\ \vdots & \vdots & \ddots & \vdots \\ z_{Nt}z_{1t} & z_{Nt}z_{2t} & \cdots & z_{Nt}^2 \end{bmatrix}. \] (12.7)

A special case of the DCC model is the constant correlation model that arises when \( \alpha = \beta = 0 \) resulting in \( Q_t = \overline{Q} \) and \( R_t = \overline{R} \). Note that there are no tests of constancy of correlations directly against the DCC model because the DCC model is only identified if correlations are changing (see Silvennoinen and Teräsvirta, 2009).

An alternative specification of the evolution of the correlation matrix, \( R_t \), is provided by the varying correlation model (VCC) of Tse and Tsui (2002). In this model the dynamics are specified directly for the matrix \( R_t \) and are given by

\[ R_t = (1 - \alpha - \beta)S + \alpha S_{t-1} + \beta R_{t-1}, \]

where \( S \) is symmetric positive definite matrix with ones on the main diagonal and \( S_{t-1} \) is a sample correlation matrix of the past \( M \) standardised residuals. The typical element of \( S_{t-1} \) is given by

\[ s_{ij t-1} = \frac{\sum_{m=1}^{M} z_{i t-m} z_{j t-m}}{\sqrt{\left( \sum_{m=1}^{M} z_{i t-m}^2 \right) \left( \sum_{m=1}^{M} z_{j t-m}^2 \right)}}. \]

A necessary condition for \( S_{t-1} \) to be positive definite is that the size of the window \( M \) be greater than the number of assets in the system being estimated.

The estimation of the DCC model for the \( N = 2 \) asset case proceeds as follows. The stationary univariate GARCH(1, 1) models for the two assets

\[ h_{1t} = \alpha_0 + \alpha_{11} u_{1t-1}^2 + \beta_{11} h_{1t-1} \]
\[ h_{2t} = \alpha_0 + \alpha_{12} u_{2t-1}^2 + \beta_{12} h_{2t-1} \]

are used to construct the standard deviation matrix

\[ S_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 \\ 0 & \sqrt{h_{2t}} \end{bmatrix}. \]
12.4. MULTIVARIATE CONDITIONAL CORRELATION

Compute the standardised residuals
\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} = S_t^{-1} u_t = \begin{bmatrix}
\frac{1}{\sqrt{h_{11t}}} & 0 \\
0 & \frac{1}{\sqrt{h_{22t}}}
\end{bmatrix} \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix} = \begin{bmatrix}
\frac{u_{1t}}{\sqrt{h_{11t}}} \\
\frac{u_{2t}}{\sqrt{h_{22t}}}
\end{bmatrix},
\]
and \( \overline{Q} \)
\[
\overline{Q} = \begin{bmatrix}
\overline{q}_{11} & \overline{q}_{12} \\
\overline{q}_{12} & \overline{q}_{22}
\end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix}
z_{1t}^2 & z_{1t} z_{2t} \\
z_{2t} z_{1t} & z_{2t}^2
\end{bmatrix}.
\]

The matrix \( Q_t \) is now given by
\[
\begin{align*}
Q_t &= (1 - \alpha - \beta) \begin{bmatrix}
\overline{q}_{11} & \overline{q}_{12} \\
\overline{q}_{12} & \overline{q}_{22}
\end{bmatrix} \\
&+ \alpha \begin{bmatrix}
z_{1t-1}^2 & z_{1t-1} z_{2t-1} \\
z_{2t-1} z_{1t-1} & z_{2t-1}^2
\end{bmatrix} + \beta \begin{bmatrix}
q_{11t-1} & q_{12t-1} \\
q_{12t-1} & q_{22t-1}
\end{bmatrix},
\end{align*}
\]
and the correlation matrix is
\[
R_t = \begin{bmatrix}
1 & \frac{q_{12t}}{\sqrt{q_{11t} q_{22t}}} \\
\frac{q_{12t}}{\sqrt{q_{11t} q_{22t}}} & 1
\end{bmatrix} = \begin{bmatrix}
1 & \rho_{12t} \\
\rho_{12t} & 1
\end{bmatrix}.
\]

The final step is the construction of the conditional covariance matrix, which is given by
\[
H_t = \begin{bmatrix}
\sqrt{h_{11t}} & 0 \\
0 & \sqrt{h_{22t}}
\end{bmatrix} \begin{bmatrix}
1 & \frac{q_{12t}}{\sqrt{q_{11t} q_{22t}}} \\
\frac{q_{12t}}{\sqrt{q_{11t} q_{22t}}} & 1
\end{bmatrix} \begin{bmatrix}
\sqrt{h_{11t}} & 0 \\
0 & \sqrt{h_{22t}}
\end{bmatrix} = \begin{bmatrix}
h_{11t} & \frac{q_{12t}}{\sqrt{q_{11t} q_{22t}}} \sqrt{h_{11t} h_{22t}} \\
\frac{q_{12t}}{\sqrt{q_{11t} q_{22t}}} \sqrt{h_{11t} h_{22t}} & h_{22t}
\end{bmatrix}.
\]

Notice that the covariance (off-diagonal terms of \( H_t \) are identical) is now time-varying as a result of the two GARCH models \( h_{11t} \) and \( h_{22t} \) and the time-varying correlation \( r_{12t} = q_{12t} / \sqrt{q_{11t} q_{22t}} \).

12.4.3 Dynamic Equicorrelation (DECO) Model

The key assumption underlying the Dynamic Equicorrelation (DECO) model of Engle and Kelly (2009) is that the unconditional correlation matrix of systems of financial returns has entries of roughly similar magnitude. The contemporaneous correlations of the DCC model are therefore assumed to be
equal across all \( N \) variables, but not over time. The pertinent restrictions on the correlation matrix are
\[
R_t = \begin{bmatrix}
1 & r_{12t} & \cdots & r_{1Nt} \\
r_{21t} & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & r_{N-1 Nt} \\
r_{N1t} & \cdots & r_{NNt-1} & 1 \\
\end{bmatrix} = \begin{bmatrix}
1 & \bar{r}_t & \cdots & \bar{r}_t \\
\bar{r}_t & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \bar{r}_t \\
\bar{r}_t & \cdots & \bar{r}_t & 1 \\
\end{bmatrix},
\]
where \( \bar{r}_t \) represents the average of the \( N(N - 1)/2 \) correlations at time \( t \)
\[
\bar{r}_t = \frac{2}{N(N - 1)} \sum_{i>j} r_{ijt}.
\]
Rewriting these restrictions shows that
\[
R_t = (1 - \bar{r}_t) \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & \cdots & 1 \\
\end{bmatrix} + \bar{r}_t \begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & 1 \\
1 & 1 & \cdots & 1 \\
\end{bmatrix} = (1 - \bar{r}_t) I_N + \bar{r}_t O_N,
\]
where \( I_N \) is the identity matrix and \( O_N \) is a \((N \times N)\) matrix of ones. This form of the correlation matrix, \( R_t \), has the following analytical properties
\[
|R_t| = (1 - \bar{r}_t)^{N-1} (1 + (N - 1) \bar{r}_t) \quad (12.8)
\]
\[
R_t^{-1} = \frac{1}{1 - \bar{r}_t} I_N - \frac{\bar{r}_t}{(1 - \bar{r}_t)(1 + (N - 1) \bar{r}_t)} O_N. \quad (12.9)
\]
As will become apparent, these expressions greatly simplify the construction of the log-likelihood function when the model is estimated.

### 12.5 Estimation

Following the methods introduced in Chapter 10, for a sample of \( t = 1, 2, \ldots, T \), observations, the log-likelihood function of a multivariate GARCH model is given by
\[
\log L = \frac{1}{T} \sum_{t=1}^{T} \log L_t = \frac{1}{T} \sum_{t=1}^{T} \log f(r_{1t}, r_{2t}, \ldots, r_{Nt})
\]
where \( f(r_{1t}, r_{2t}, \ldots, r_{Nt}) \) is a \( N \)-dimensional multivariate probability distribution. To implement the estimation by maximum likelihood methods, it is necessary to specify the functional form of the \( N \)-dimensional probability distribution. There are two popular choices in empirical applications, namely, the multivariate normal distribution and the multivariate \( t \) distribution.
12.5. Estimation

12.5.1 Log-likelihood Functions

Multivariate Normal Distribution

The multivariate normal distribution is given by

\[ f(r_1, r_2, \cdots, r_N) = \left( \frac{1}{2\pi} \right)^{N/2} |H_t|^{-1/2} \exp \left( -0.5 u_t' H_t^{-1} u_t \right) \]

where \(u_t\) is the \((N \times 1)\) vector of disturbances at time \(t\). The disturbances are obtained from

\[
\begin{bmatrix}
  u_{1t} \\
  u_{2t} \\
  \vdots \\
  u_{N,t}
\end{bmatrix}
= 
\begin{bmatrix}
  r_{1t} \\
  r_{2t} \\
  \vdots \\
  r_{N,t}
\end{bmatrix}
- 
\begin{bmatrix}
  \mu_{1t} \\
  \mu_{2t} \\
  \vdots \\
  \mu_{N,t}
\end{bmatrix}
\]

in which \(r_t\) as a \((N \times 1)\) vector of returns, \(\mu_t\) is the \((N \times 1)\) vector of conditional means of the returns series and \(H_t\) is the conditional covariance matrix of the returns.

In the case of the multivariate conditional normality, the log-likelihood function at time \(t\) takes the form

\[
\log L_t = \log f(r_1, r_2, \cdots, r_N)
= -0.5N \log (2\pi) - 0.5 \log |H_t| - 0.5u_t' H_t^{-1} u_t. \tag{12.10}
\]

Multivariate (Standardised) Student t Distribution

The standardised multivariate \(t\) distribution is given by

\[
f(r_1, r_2, \cdots, r_N) = \frac{\Gamma \left( \frac{\nu+N}{2} \right)}{(\pi (\nu-2))^{N/2} \Gamma \left( \frac{\nu}{2} \right)} \times |H_t|^{-1/2} \left( 1 + \frac{u_t' H_t^{-1} u_t}{\nu - 2} \right)^{-(\nu+N)/2}.
\]

As for case of the normal distribution,

\[u_t = r_t - \mu_t\]

in which \(r_t\) as a \((N \times 1)\) vector of returns, \(\mu_t\) as a \((N \times 1)\) vector of conditional means and \(H_t\) is the conditional covariance matrix. The parameter \(\nu\) is the degrees of freedom parameter whereby small values represent fat-tails.
In this case, the log-likelihood function at time $t$ is

$$
\log L_t = \log f(r_{1t}, r_{2t}, \ldots, r_{Nt}) = \log \Gamma \left( \frac{\nu + N}{2} \right) - \frac{N}{2} \log \left( \pi \left( \frac{\nu - 2}{2} \right) \right) - \log \Gamma \left( \frac{\nu}{2} \right) - \frac{1}{2} \log |H_t| - \left( \frac{\nu + N}{2} \right) \left( 1 + \frac{u_t H_t^{-1} u_t}{\nu - 2} \right)
$$

(12.11)

The log-likelihood function in (12.10) or (12.11) is then maximised using standard optimisation algorithms. In the case of VECH and BEKK models in Section 12.3 the matrix $H_t$ is constructed directly. For the correlation models in Section 12.4.1 $H_t$ is constructed as either as $H_t = S_t R S_t$ (constant correlations) or $H_t = S_t R_t S_t$ (dynamic correlations). The advantage of the results in equations (12.8) and (12.9) for the DECO model are now clear because the matrix computations in the log-likelihood function requiring the determinant and the inversion of $H_t$ are greatly simplified.

### 12.5.2 Estimating the BEKK Model

Consider estimating a bivariate diagonal BEKK model for the percentage excess returns to Microsoft, $r_{1t}$, and the S&P 500 index, $r_{mt}$, used in Section 12.1. The model is given by

$$
\begin{align*}
& r_{1t} = \gamma_1 + u_{1t} \\
& r_{mt} = \gamma_2 + u_{2t}
\end{align*}
$$

with

$$
\begin{align*}
u_t &= \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim N(0, H_t) \\
H_t &= \begin{bmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{bmatrix} = CC' + Au_{t-1} u_{t-1}' A' + BH_{t-1} B' \\
C &= \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}.
\end{align*}
$$

The maximum likelihood parameter estimates of the bivariate BEKK model using a log-likelihood function based on the normal distribution are

$$
\begin{align*}
& r_{1t} = 2.170 + \hat{u}_{1t} \\
& r_{mt} = 0.454 + \hat{u}_{2t},
\end{align*}
$$
with estimated conditional covariance matrix, $\hat{H}_t$, given by

\[
\begin{bmatrix}
\hat{h}_{11t} & \hat{h}_{12t} \\
\hat{h}_{12t} & \hat{h}_{22t}
\end{bmatrix}
= \begin{bmatrix}
9.784 & 0.000 \\
1.279 & 0.795
\end{bmatrix}
+ \begin{bmatrix}
0.365 & 0.000 \\
0.000 & 0.248
\end{bmatrix}
\begin{bmatrix}
\hat{u}_{1t-1}^2 & \hat{u}_{1t-1}\hat{u}_{2t-1} \\
\hat{u}_{2t-1}\hat{u}_{1t-1} & \hat{u}_{2t-1}^2
\end{bmatrix}
\begin{bmatrix}
0.365 & 0.000 \\
0.000 & 0.248
\end{bmatrix}
+ \begin{bmatrix}
0.875 & 0.000 \\
0.000 & 0.941
\end{bmatrix}
\begin{bmatrix}
\hat{h}_{11t-1} & \hat{h}_{12t-1} \\
\hat{h}_{12t-1} & \hat{h}_{22t-1}
\end{bmatrix}
\begin{bmatrix}
0.875 & 0.000 \\
0.000 & 0.941
\end{bmatrix}
\]

Figure 12.4: Estimates of the time-varying covariance and associated beta estimated using a diagonal BEKK model for the percentage excess returns to Microsoft and the S&P 500 index. The data are monthly for the period April 1990 to July 2004 ($T = 172$).

Figure 12.4 plots the conditional covariance between the excess returns on the market and Microsoft and the associated time-varying estimate of beta risk. These results should be compared with those in Figures 12.1 and 12.2 in Section 12.1, which were generated using the constant correlation assumption. The BEKK results produce a much steeper fall in the conditional covariance at the beginning of the sample than do the earlier constant correlation results. This major difference causes the time-variation in beta risk to be quite different in the first half of the sample period. However, during the second half of the period, the effect of the DotCom bubble is shown in the sharp increase in beta risk of Microsoft and this effect is common to both sets of results.

The early multivariate GARCH models are known to have their problems: the VECM model suffers because of the dimension of the parameter space; the BEKK model has parameters which are difficult to interpret. It is probably fair to say, therefore, that recent empirical work has concentrated more on the multivariate correlation class of models.
12.5.3 Estimating Multivariate Correlation Models

Maximum likelihood estimation of this class of model is illustrated using daily returns data on United States industry portfolios for the period 1 January 1990 to 31 December 2008. The industries considered are: Consumer Durables, NonDurables, Wholesale, Retail, and Services (Csnmr); Manufacturing, Energy, and Utilities (Manuf); Business Equipment, Telephone and Television Transmission (HiTec); and Healthcare, Medical Equipment, and Drugs (Hlth).

The returns to the industry portfolios are plotted in Figure 12.5. It is immediately apparent that all the stocks experience an increase in volatility at the end of the sample period as the global financial crisis begins. There is also evidence that the DotCom bubble had a much greater influence on the volatility of the technology industry than on the others, an observation that confirms the potential advantages of estimating multivariate GARCH models allow for volatility spillovers from one industry to another. To provide a comparison of the correlation models discussed in Section 12.4.1, the CCC and DCC models will be estimated using likelihood functions based on both the multivariate normal and multivariate standardised Student $t$ distribution. Although the estimation of the multivariate correlation models appears quite
complex, reasonable starting values for the optimisation can be obtained in a systematic approach as follows.

1. The starting values for the parameters in the mean equations can be obtained by simple ordinary least squares regression.

2. The starting values for the parameters in the variance equations can be found by fitting univariate GARCH models to each of the series.

3. The estimates of the conditional variances can be used to standardise the residuals and obtain starting values for correlations.

4. This leaves only the parameters $\alpha$ and $\beta$ the govern the dynamics of the correlation to be provided. These parameters must satisfy the constraints $\alpha, \beta \geq 0$ and $0 \leq \alpha + \beta < 1$, so a fairly safe guess at starting values would be to choose values for these parameters that are similar to values obtained from univariate GARCH models with $\alpha$ taken to be of the order of the coefficient on the lagged squared residuals (news) and $\beta$ being of the order of the coefficient on the lagged conditional variance. A more formal procedure would be to perform a crude two-dimensional grid search.

Once these starting parameters are provided, the log-likelihood function is optimised using one a standard iterative optimisation algorithm.

The parameter estimates for the various models are reported in Table 12.3.

The sums of the parameters on the ARCH, $\alpha_{ii}$, and GARCH, $\beta_{ii}$, terms in the conditional variance part of models are all close to unity and indeed in a couple of instances marginally exceeds unity. This indicates that the volatility of these portfolios is fairly persistent. The adjustment parameters $\alpha$ and $\beta$ on the DCC model are statistically significant and this suggests that there is statistical evidence to support the claim that correlations are time varying, despite the fact note previously that it is difficult to test this hypothesis formally. There is some scope however for exploring the hypothesis of constant correlations using a residual-based method suggested by Bollerslev (1990). If the multivariate GARCH model in equation (12.1) is correctly specified then conditional on information set at $t - 1$, it follows that $E(u_{it}u_{jt}) = h_{ijt}$. Bollerslev (1990) suggests pairwise residual based regression diagnostic tests for the adequacy of a constant correlation specification. The procedure is as follows:

\[ i = j: \ \text{Regess} \frac{\hat{u}_{it}^2}{h_{ijt}} \text{ on } \frac{1}{h_{ijt}}, \frac{\hat{u}_{it-1}^2}{h_{ijt}}, \ldots, \frac{\hat{u}_{it-q}^2}{h_{ijt}} \]

\[ i \neq j: \ \text{Regess} \frac{\hat{u}_{it}\hat{u}_{jt}}{h_{ijt}} \text{ on } \frac{1}{h_{ijt}}, \frac{\hat{u}_{it-1}^2}{h_{ijt}}, \frac{\hat{u}_{jt-1}^2}{h_{ijt}}, \frac{\hat{u}_{it-1}\hat{u}_{jt-1}}{h_{ijt}}, \ldots, \frac{\hat{u}_{it-q}\hat{u}_{jt-q}}{h_{ijt}}, \]

and test the null hypothesis that all the regressors are zero using a conventional $F$ test.
Coefficient estimates for the CCC and DCCC models for likelihood functions based on the multivariate normal and the multivariate standardised student $t$ distribution. The data are daily returns for 4 industry portfolios (Cnsmr=1, Manuf=2, Hitec=3, Hlth=4). The sample period is 1 January 1990 to 31 December 2008. The parameter $\nu$ represents the degrees of freedom parameter when estimation is based on the $t$ distribution. Standard errors are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Normal Distribution</th>
<th>Student $t$ Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ccc</td>
<td>dcc</td>
</tr>
<tr>
<td>$\mu_{01}$</td>
<td>0.054</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>$\alpha_{01}$</td>
<td>0.005</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.041</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.955</td>
<td>0.936</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\mu_{02}$</td>
<td>0.063</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>$\alpha_{02}$</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.039</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.959</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\mu_{03}$</td>
<td>0.063</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>$\alpha_{03}$</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\alpha_{13}$</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>0.958</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\mu_{04}$</td>
<td>0.060</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>$\alpha_{04}$</td>
<td>0.005</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\alpha_{14}$</td>
<td>0.039</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>0.958</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>6.605</td>
<td>9.089</td>
</tr>
<tr>
<td></td>
<td>(0.298)</td>
<td>(0.499)</td>
</tr>
</tbody>
</table>
These pairwise test statistics cannot provide evidence in favour of a DCC specification over a the CCC model, neither can they distinguish between misspecification due to the incorrect treatment of correlations as opposed to the use of an incorrect GARCH model for the conditional covariances. However, the tests do shed some light on the adequacy of the assumption of constant correlations and also their pairwise construction helps to gain an understanding as to where any problems with the constant correlation assumption are to be found. In the case of the industry portfolios, when \( i = j \) the \( F \) statistics are 2.127, 3.324, 2.784 2.785 which all reject the null hypothesis (although for the consumer portfolio the rejection is only at the 10% level). These results are not as interesting as the \( i \neq j \) cases and in particular the tests involving the portfolio of high technology stocks. The \( F \) tests for constant correlations between high technology and consumer goods, high technology and manufacturing and high technology and health, are 216.642, 7.486 and 94.915, respectively. These test values are orders of magnitude higher than any other recorded test statistic. This indicates very strongly that high technology stocks behaved differently to the others over the sample period in the context of the constant correlation assumption. Of course this behaviour is clearly seen in Figure 12.5 where the volatility of high technology stocks during the DotCom crisis is clearly different to those of the other industries. Once again, these results should be regarded merely as preliminary evidence and certainly does not provide an argument in favour of using the DCCC model. Indeed testing either the assumption of dynamic conditional correlation model or even the adequacy of the constant correlation assumption is a matter of ongoing research (Harvey and Thiele, 2015; Silvennoinen and Teräsvirta, 2015).

The symmetric quasi-correlation matrix for the DCC model in equation (12.7), for estimation based on the normal distribution is given by

\[
\begin{bmatrix}
1 & 0.7948 & \cdots \\
0.7948 & 1 & \cdots \\
\vdots & \vdots & \ddots \\
\end{bmatrix}
\]

The standard errors shown in parentheses are obtained using the delta method. It is fairly clear from these results that the quasi-correlations would not support the restriction that they are all equal and hence the DECO model is not indicated in this instance.

There is also an argument to be made in favour of the multivariate \( t \) distribution over the normal distribution as the basis for the construction of the log-likelihood function. The degrees of freedom parameter of the \( t \) distribution, \( \nu \), is estimated quite precisely in both the CCC and DCC models. Once again a formal test of this hypothesis must be carefully formulated as the null hypothesis is not \( \nu = 0 \) as in the traditional \( t \) test, neither is it \( \nu = \infty \) which would be the case if a normal distribution were the appropriate choice. In
practice, the normal distribution and the $t$ distribution become indistinguishable for $\nu > 30$ and it is quite clear from the point estimates $\hat{\nu}$ and the estimated standard errors that the hypothesis $\nu = 30$ would be strongly rejected.

### 12.6 Capital Ratios and Financial Crises

The global financial crisis of 2007–2009 heightened awareness of the importance of systemic risk to financial policymaking. Broadly speaking the systemic risk of a financial institution is the contribution the institution would make to the deterioration of the entire financial system in a crisis period. In a recent paper, Brownlees and Engle (2012) introduce a method for determining the marginal expected shortfall (MES) of a financial institution, defined as the expected loss an investor in the financial firm would experience if the market declined substantially. In order to estimate MES, a multivariate volatility model must be specified, estimated and simulated in order to provide a view of the potential evolution of market and firm returns.

The actual capital ratio of a firm at time $t$ is defined as

$$\text{ACF}_{it} = \frac{W_{it}}{W_{it} + D_{it}}, \quad (12.12)$$

where $W_{it}$ is the firm’s equity value, $D_{it}$ is the book value of its debt and $W_{it} + D_{it}$ represents the total value of the firm’s assets. Table 12.4 gives the actual capital ratios of 18 financial institutions in the United States at the end of 2014, with the institutions sorted in terms of their equity values, $W_{it}$. The important question is whether these capital ratios are sufficient for the firms to remain solvent during periods of financial distress in the future. To tackle this question, Brownlees and Engle (2012) derive a safe measure of capital using a multivariate GARCH model of asset returns.

To derive the safe capital ratio, define the working capital of a firm at time $t$ to be the difference between its equity value $W_t$ and a proportion $k$ of its total assets $W_t + D_t$,

$$K_{it} = W_{it} - k(W_{it} + D_{it}). \quad (12.13)$$

During a financial crisis there is a large fall in the market return, $r_{mt+h}$, in excess of some threshold, $c$, which can result in a capital shortfall, $CS$, in the future given by

$$CS_{it+h} = -E(K_{it+h} | r_{mt+h} < c). \quad (12.14)$$

Using (12.13) the capital shortfall is rewritten as

$$CS_{it+h} = -E(W_{it+h} - k(W_{it+h} + D_{it+h}) | r_{mt+h} < c)$$

$$= E(D_{it+h} | r_{mt+h} < c) - (1 - k) E(W_{it+h} | r_{mt+h} < c). \quad (12.15)$$

Assuming that debt cannot be renegotiated $E(D_{it+h} | r_{mt+h} < c) = D_{it}$ and using the result that the future equity value of the firm is $W_{it+h} = W_{it}(1 -$
12.6. CAPITAL RATIOS AND FINANCIAL CRISES

Table 12.4

Capital ratios of selected financial institutions as at 31 December 2014. Actual capital ratios are defined as in equation (12.12).

<table>
<thead>
<tr>
<th>Institution</th>
<th>Equity Value (E)</th>
<th>Actual Capital Ratios (ACR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wells Fargo</td>
<td>284385548</td>
<td>0.163666</td>
</tr>
<tr>
<td>JPMorgan Chase</td>
<td>233935868</td>
<td>0.092507</td>
</tr>
<tr>
<td>Bank of America</td>
<td>188139291</td>
<td>0.090744</td>
</tr>
<tr>
<td>Citigroup</td>
<td>163925596</td>
<td>0.089366</td>
</tr>
<tr>
<td>American Express</td>
<td>96266348</td>
<td>0.418410</td>
</tr>
<tr>
<td>Goldman Sachs Group</td>
<td>84421881</td>
<td>0.097466</td>
</tr>
<tr>
<td>Morgan Stanley</td>
<td>75947236</td>
<td>0.092937</td>
</tr>
<tr>
<td>Capital One Financial</td>
<td>45895407</td>
<td>0.151976</td>
</tr>
<tr>
<td>Bank of New York Mellon</td>
<td>45670055</td>
<td>0.116009</td>
</tr>
<tr>
<td>State Street</td>
<td>32773358</td>
<td>0.114548</td>
</tr>
<tr>
<td>SunTrust Banks</td>
<td>21849006</td>
<td>0.117233</td>
</tr>
<tr>
<td>Fifth Third Bank</td>
<td>16789143</td>
<td>0.123762</td>
</tr>
<tr>
<td>Northern Trust</td>
<td>15873037</td>
<td>0.134048</td>
</tr>
<tr>
<td>Regions Financial</td>
<td>14535576</td>
<td>0.124688</td>
</tr>
<tr>
<td>KeyCorp</td>
<td>12041918</td>
<td>0.131752</td>
</tr>
<tr>
<td>Huntington Bank</td>
<td>8568056</td>
<td>0.128700</td>
</tr>
<tr>
<td>Comerica</td>
<td>8416539</td>
<td>0.120482</td>
</tr>
<tr>
<td>Zions Bank</td>
<td>5785591</td>
<td>0.107296</td>
</tr>
</tbody>
</table>

\[ r_{it+h} \], where \( r_{it+h} \) is the future return on the firm, this expression is rewritten as

\[
CS_{it+h} = D_{it} - (1 - k)W_{it}E((1 - r_{it+h})|r_{mt+h} < c)
\]

\[
= D_{it} - (1 - k)W_{it}(1 - \text{MES}_{it}), \quad (12.16)
\]

where \( \text{MES}_{it+h} = E(r_{it+h}|r_{mt+h} < c) \), (12.17)

represents the marginal expected shortfall.

The safe capital ratio is defined as that ratio for which it is not necessary to raise any additional external capital during a crisis. By setting \( CS_{it+h} = 0 \) in (12.15) and rearranging, the safe capital ratio for the firm at time \( t \) is a function of the marginal expected shortfall, \( \text{MES} \), and the prudential parameter \( k \) given by

\[
\text{SCR}_{it} = \frac{k}{1 - (1 - k)\text{MES}_{it+h}}. \quad (12.18)
\]

Consider estimating the \( h = 1 \) day marginal expected shortfall in (12.17) for Morgan Stanley, \( r_{1t} \), where the market returns, \( r_{mt} \), given by the value
weighted returns of the S&P 500 index. The first step is to estimate the bivariate CCC model

\[ r_m t = \gamma_m + u_m t \]
\[ r_i t = \gamma_i + u_i t, \quad (12.19) \]

where

\[ u_t = \begin{bmatrix} u_m t \\ u_i t \end{bmatrix} \sim N(0, H_t) \]

and the conditional variance matrix is

\[ H_t = S_t R S_t = \begin{bmatrix} \sqrt{h_m t} & 0 \\ 0 & \sqrt{h_i t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{m1} \\ \rho_{m1} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_m t} \\ 0 \end{bmatrix} \quad (12.20) \]

in which \( \rho_{m1} \) is the constant correlation between market excess returns and Morgan Stanley excess returns. The conditional variances, \( h_m t \) and \( h_i t \), are generated from univariate GARCH(1,1) models

\[ h_m t = \alpha_0 + \alpha_1 u_m^2 t - 1 + \beta_1 h_m t - 1 \]
\[ h_i t = \alpha_0 + \alpha_1 u_i^2 t - 1 + \beta_1 h_i t - 1. \quad (12.21) \]

The parameter estimates of the bivariate CCC model in equations (12.19), (12.20) and (12.21) using daily returns data for the period 14 December 2001 to 31 December 2014 are given in Table 12.5. The GARCH parameter estimates for the two returns display typical empirical features of near integrated conditional volatility models, with the estimates on the lagged squared residuals around 0.1 and the estimates of the lagged conditional variances around 0.9, with the sum of the parameter estimates for each asset being close to unity. The estimated correlation coefficient is \( \hat{\rho}_{mi} = 0.7290. \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Market</th>
<th>Morgan Stanley</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_i )</td>
<td>0.069366</td>
<td>0.0143008</td>
</tr>
<tr>
<td>( \alpha_{0i} )</td>
<td>0.0222742</td>
<td>0.0033543</td>
</tr>
<tr>
<td>( \alpha_{1i} )</td>
<td>0.0775348</td>
<td>0.0073557</td>
</tr>
<tr>
<td>( \beta_{1i} )</td>
<td>0.899839</td>
<td>0.0090589</td>
</tr>
<tr>
<td>( \rho_{1m} )</td>
<td>0.7290</td>
<td>0.0082132</td>
</tr>
</tbody>
</table>

The second step is to estimate the 1-day MES given the conditional variance estimates. The approach is to follow Brownlees and Engle (2012) by using
12.7. OPTIMAL HEDGE RATIOS

stochastic simulation methods. This involves simulating the estimated bi
variate CCC model in (12.19) and (12.20) where \( u_t \) is drawn with replacement from the estimated residuals \( \hat{u}_t \). This is repeated \( S \) times which generates \( s = 1, 2, \ldots, S \) simulated returns \( r^s_{mt} \) and \( r^s_{it} \). The estimated MES is then computed by extracting the simulated values \( r^s_{mt} \) for those cases where the simulated market returns \( r^s_{mt} \) are less than \( c \). Formally, the \( h = 1 \) day MES is computed as

\[
MES = \frac{\sum_{s=1}^{S} r^s_{it} I (r^s_{mt} < c)}{\sum_{s=1}^{S} I (r^s_{mt} < c)},
\]

(12.22)

where \( I (r^s_{mt} < c) \) is an indicator function defined as

\[
I (r^s_{mt} < c) = \begin{cases} 
1 & : \ r^s_{mt} < c \\
0 & : \ r^s_{mt} \geq c.
\end{cases}
\]

(12.23)

The estimated 1-day MES for Morgan Stanley with \( c = -0.02 \) and \( S = 10000 \) simulations, is \( MES = 0.056 \). The safe capital ratio with a prudential parameter of \( k = 0.08 \) is

\[
SCR = \frac{k}{1 - (1 - k)MES} = \frac{0.02}{1 - (1 - 0.02)MES} = 0.013.
\]

This value is much smaller than the actual capital ratio of \( ACR = 0.092937 \) for this company given in Table 12.4. This of course accords with intuition as it is to be expected that the banks would be well situated to deal with any MES over a 1-day horizon. Of course, the real question of the adequacy of the capital structure of Morgan Stanley can only be gauged under simulation if much longer time horizons are used.

12.7 Optimal Hedge Ratios

Another important empirical application of multivariate GARCH models has been in the area of the estimation of dynamic hedge ratios (Baillie and Myers, 1991). Consider an investor who sells futures contracts to hedge against movements in the spot price of an asset rate. The return on the hedged portfolio, \( r_{ht} \), is

\[
r_{ht} = r_{st} - \eta r_{ft}
\]

where \( r_{st} \) is the return in the spot market, \( r_{ft} \) is the return on the futures contract and \( \eta \) is the number of contracts the hedger sells for each unit of spot commodity, known as the hedge ratio. The expected return on the portfolio is

\[
\mu_h = E[r_{s,t} - \eta r_{f,t}] = E[r_{s,t}] - \eta E[r_{f,t}] \quad \text{or} \quad \mu_h = \mu_s - \eta \mu_f,
\]
and the variance of the portfolio is

\[
\sigma^2_h = E[(r_{ht} - \mu_h)^2] = E[((r_{st} - \mu_s) - \eta(r_{ft} - \mu_f))^2]
\]

\[
= E[(r_{st} - \mu_s)^2] + \eta^2 E[(r_{ft} - \mu_f)^2]
- 2\eta E[(r_{st} - \mu_s)(r_{ft} - \mu_f)]
\]

\[
= \sigma^2_s + \eta^2 \sigma^2_f - 2\eta \sigma_{sf}.
\] (12.24)

The optimal minimum variance portfolio is found by minimising \(\sigma^2_h\) by choice of \(\eta\). Differentiating expression (12.24) with respect to \(\eta\) gives

\[
\frac{d\sigma^2_h}{d\eta} = 2\eta \sigma^2_f - 2\sigma_{sf}.
\]

Setting this derivative to zero and solving for \(\eta\) gives the optimal hedge ratio

\[
\eta = \frac{\sigma_{sf}}{\sigma^2_f},
\] (12.25)

which is the ratio of the covariance of the returns on the spot and futures contracts to the variance of the return on futures. The objective of variance minimisation assumes a high degree of risk aversion on the part of economic agents. However, Baillie and Myers (1989) show that if the expected returns to holding futures are zero, the minimum variance hedging rule is also the expected utility-maximising rule.

The expression for the optimal hedge ratio in (12.25) assumes that the covariance and the variance are constant. This, in turn, results in a hedge ratio that is also constant, implying that the hedger never rebalances the portfolio in response to shocks in the spot and futures markets. To relax the restriction of a constant hedge ratio, the covariance of the returns on the spot and futures contracts and the variance of the return on futures contract must be specified as time-varying. The resultant dynamic hedge ratio is then

\[
\eta_t = \frac{\sigma_{sf_t}}{\sigma^2_{ft_t}}
\]

which is now the time-varying ratio of the conditional covariance of the returns on the spot and futures contracts to the conditional variance of the return on futures. To model the time-variation in the conditional covariance and variance, a bivariate GARCH model is required.

Consider the problem of hedging the returns to the 4 United States industry portfolios using the daily returns for the period is 1 January 1990 to 31 December 2008 using the futures contract on the S&P 500 index for the same period. The constant hedge ratio for each of the industry portfolios is found by estimating the regression

\[
r_{jt} = \beta_{j0} + \beta_{j1}r_{ft} + u_{jt}, \quad j = 1, 2, 3, 4,
\]
with \( r_{jst} \) representing the returns to the relevant industry portfolio and \( r_{f1} \) represents the returns to the three month S&P 500 futures contract. The constant hedge ratios are estimated to be 
\[
\text{Cnsmr} = 0.773, \quad \text{Manuf} = 0.787, \quad \text{Hitec} = 1.124, \quad \text{Hlth} = 0.768.
\]

**Figure 12.6:** Dynamic hedge ratios (solid line) for 4 United States industry portfolios hedged using the 3-month S&P index futures contract. The relevant time-varying variances and covariances are computed using a DCC model using daily data for the period 1 January 1990 to 31 December 2008. The optimal constant hedge ratio is shown as dashed line.

The dynamic hedge ratios are computed using bivariate dynamic conditional correlation (DCC) models specified for each industry portfolio relative to the S&amp;P500 futures contract. The model is estimated using the normal distribution. The parameter estimates for the bivariate are not reported but are not vastly different from the multivariate DCC models reported in Table 12.3. The dynamic hedge ratios computed in this manner for each industry portfolio are plotted in Figure 12.6.

For the consumer goods industry the constant hedge ratio looks like a reasonable strategy and the value of the dynamic ratio is seldom strays too far away from this constant value. The same conclusion cannot be made about the other three portfolios and particularly the manufacturing and high technology industries. The dynamic hedge ratio for manufacturing is below the
constant value for most of the early part of the sample and then switches to being above it after the DotCom bubble unwinds in the early 2000s. The effect of the DotCom bubble on high technology stocks is very clear and the advantage of using dynamic hedging is obvious. Finally, while the dynamic hedge ratio for the health portfolio fluctuates around the constant hedge ratio, the deviations during the early 1990s (above the constant ratio) and the DotCom bubble (below the line) suggest that dynamic hedging would provide a substantial reduction in risk exposure.

12.8 Exercises

1. Bivariate Constant Correlation Models

| capm.csv, capm.dta, capm.wf1 |

The data are monthly observations for the period April 1990 to July 2004 ($T = 172$) on the share prices of five United States stocks and also the price of the commodity gold as well as the S&P 500 index.

(a) Compute the excess returns to the market portfolio (S&P 500 index) and the excess returns to Microsoft. Estimate Microsoft’s constant beta risk using the CAPM model.

(b) Estimate a GARCH(1,1) model for the excess returns to the market and excess returns to Microsoft, respectively. Comment on your results.

(c) Based on the assumption of a constant correlation between Microsoft and the market, compute an estimate of the conditional covariance between Microsoft and the market using the conditional variances obtained in (b). Hence provide an estimate of time-varying beta risk for Microsoft.

(d) Is the estimate of the time-varying beta risk significantly affected by the introduction of a leverage effect in the univariate GARCH models?

(e) Is the estimate of time-varying beta risk significantly affected by the use of a multivariate diagonal VECH or BEKK model?

(f) Compute the monthly excess returns to Walmart and estimate a GARCH(1,1) model using these returns. Based on the assumption of constant correlation between Microsoft and Walmart compute the optimal time-varying portfolio weights for this two asset portfolio. Contrast your results with the optimal constant portfolio weights.

(g) Now estimate a multivariate GARCH model (either a diagonal VECH or BEKK) model for Microsoft and Walmart and re-compute
the optimal weights. Are the results significantly different to those obtained in (d)?

2. Time-varying Correlation in Hedge Funds

The data are daily returns to various hedge funds for period 1 April 2003 to 28 May 2010 ($T = 1869$) obtained from Hedge Fund Research, Inc. ("HFR").

(a) Estimate a bivariate DCC model for Merger hedge fund returns and returns to the S&P500 index. Compute and plot an estimate of the time varying correlation between the Merger fund returns and the market returns. Comment on your result.

(b) Repeat part (a) for the other six hedge funds. Discuss how successful the hedge funds were in minimising exposure to systematic risk from the market during the global financial crisis from mid 2007 to the end of 2009?

(c) Now estimate a DCC model which deals with all 7 hedge fund returns (Convertible, Distressed, Equity Event, Macro, Merger and Neutral). Comment on whether or not a DECO model would be appropriate for this system.

3. Industry Portfolios

The data are daily returns on United States industry portfolios for the period 1 January 1990 to 31 December 2008. The industries considered are: Consumer Durables, NonDurables, Wholesale, Retail, and Services (Csnmr); Manufacturing, Energy, and Utilities (Manuf); Business Equipment, Telephone and Television Transmission (HiTec); and Healthcare, Medical Equipment, and Drugs (Hlth).

(a) Plot the returns to the 4 industry portfolios and comment on their time series properties.

(b) For a system comprising the 4 industry portfolio returns estimate the parameters of CCC, DCC and DECO specifications using log-likelihood functions based on the multivariate normal and the multivariate $t$ distributions, respectively. Comment on the results.
(c) Estimate the linear regression

\[
    r_t = \alpha + \beta r_{mt} + u_t
\]

where \( r_t \) is the return to the Consumer industry portfolio and \( r_{mt} \) is the return to the three month S&P futures index. What interpretation can be given to \( \hat{\beta} \) in this particular case.

(d) Estimate a bivariate DCC model for the returns to the Consumer industry portfolio and three month S&P futures index returns. Use the results of the DCC model to provide an estimate of a dynamic hedge ratio for the consumer portfolio. Interpret your results.

(e) Repeat parts (c) and (d) for the remaining three industry portfolios and comment on your results.